University of Alberta

Challenges in the application of ^{1}H-NMR spectroscopy to human metabolomic studies: lessons from asthma

by

Erik Jarl Saude

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Biochemistry

Edmonton, Alberta
Spring 2007

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-29736-0
Our file Notre référence
ISBN: 978-0-494-29736-0

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Dedicated to my loving parents:

Rev. Dr. David and Helen Saude

"The most erroneous stories are those we think we
know best – and therefore never scrutinize or question"
Stephen Jay Gould (1941 - 2002)

"The mere formulation of a problem is far more essential than its solution, which may be merely a matter of mathematical or experimental skills. To raise new questions, new possibilities, to regard old problems from a new angle requires creative imagination and marks real advances in science"
Albert Einstein (1879 – 1955)

ABSTRACT

Metabonomics has emerged as the latest field of research within the recent 'omics' generation. Like the predecessors, genomics and proteomics, metabonomics takes more of a global approach to biological research. Metabonomics detects and follows all the metabolites in a particular organism. A slightly more focused term, metabolomics, follows metabolites of a specific pathway, or of a particular disease. One-dimensional ^{1}H-NMR has become one of the principle tools for metabolomic investigations. NMR analysis of sampled biofluids, such as urine, revealed the rich spectral features from the hundreds of metabolites, each providing information about ongoing physiological processes in the organism.

The goal of this thesis was to take the theory and technology behind NMR spectroscopy and apply it to the growing field of metabolomics. Chapter II presents novel work that optimized NMR for the qualitative and quantitative analysis of metabolites in biofluids. The defined procedures for proper acquisition, processing, and analysis of NMR spectra provided a strong basis for later NMR quantitative analysis of biofluids, such as urine. Chapter III applied the quantitative analysis to identify indices of bacterial contamination in normal human urine. Once the source of the metabolic change was identified proper sample preparatory and storage measures were investigated and recommended.

A philosophical question arose regarding the metabolic baseline for human clinical and metabolomic research. Chapter IV follows the study of normal human metabolite variability, animal model variations, and the philosophical question 'What is normal', and how should researchers deal with the interplay of variability and biological homeostasis. The application of these earlier considerations are found in the metabolomic investigation of cystic fibrosis in a human clinical study, as well as the study of asthma in a guinea pig model (Chapters V and VI). The application of NMR for metabolomic studies of human pulmonary disease provided biochemical information regarding specific pathophysiological mechanisms occurring in humans and animals. Chapter VII provides concluding remarks for the thesis and offers suggestions for continuing the metabolic investigation as well as the philosophical questioning of pathways in human disease.

TABLE OF CONTENTS

Page

CHAPTER I – Challenges in the application of 1H-NMR spectroscopy to human metabolomic studies: lessons from asthma

I.A CHAPTER INTRODUCTION 1

I.B METABOLOMIC TECHNOLOGIES 7

  I.B.1 Mass spectrometry 7

  I.B.2 Nuclear magnetic resonance spectroscopy 8

I.C METABOLOMIC SAMPLES 10

  I.C.1 Sampled population 10

  I.C.2 Biological relevance 11

  I.C.3 Sampling invasiveness 12

  I.C.4 Ethics 14

  I.C.5 Subject compliance 14

  I.C.6 Sample type 15

  I.C.7 Sample preparation and storage 16

I.D URINE: A METABOLOMIC BIOFLUID 16

  I.D.1 Sampled population 17

  I.D.2 Biological relevance 18

  I.D.3 Sample acidity 19

  I.D.4 Sample preparation and storage 19

I.E HUMAN BIOLOGICAL HOMEOSTATIS VS. VARIATION 20

  I.E.1 Biological complexity 20

  I.E.2 Biological homeostasis 21

I.E.3 Metabolic homeostasis 23

I.F NORMAL METABOLOMIC POPULATIONS 24

I.F.1 Importance of normal 24

I.F.2 Definition of normal 25

I.G BIOMARKER DISCOVERY AND METABOLOIMC CORRELATION 27

I.G.1 Systems biology, complex questions 28

I.G.2 Metabolite or marker? 29

I.G.3 Population separation 30

I.G.4 Biochemical interpretation 32

I.H THESIS RATIONALE 33

Specific Aim # 1 – Optimization of NMR data acquisition for metabolomic investigations 34

Specific Aim # 2 – Urine sample preparation and storage 34

Specific Aim # 3 – Normal population 35

Specific Aim # 4 – Metabolomic analysis of sputum from CF patients 35

Specific Aim # 5 – Guinea pig model of asthma 35

Specific Aim # 6 – Analysis of acute and post-treatment human asthma 36

Specific Aim # 7 – Pediatric asthma longitudinal study 36

Specific Aim # 8 – Recommendations for future metabolomic studies 37

LITERATURE CITED 38

CHAPTER II – Optimization of NMR qualitative and quantitative analysis of body fluids

OVERVIEW 47

INTRODUCTION 48

EXPERIMENTAL PROCEDURES

II-A. Sample preparation 49

II-B. NMR spectral acquisition 50

II-C. NMR quantification 53

II-D. T_1 measurement 53

II-E. Filter attenuation mapping 54

RESULTS and DISCUSSION

II-A. Pulse sequence choice and optimization 55

II-B. Influence of relaxation 60

II-C. Filter attenuation 67

II-D. Amino acid analysis 74

II-E. Software 74

II-F. Precision 76

CONCLUSION 77

LITERATURE CITED 79

CHAPTER III – Urine as a biofluid for metabolomic studies

OVERVIEW 82

INTRODUCTION 83

EXPERIMENTAL PROCEDURES

III-A. Urine collection 85

III-B. Sample preparation 85

III-C. Sample storage 86

III-D. NMR analysis 87

III-E. NMR quantification 87

III-F. Statistical analysis 89

RESULTS and DISCUSSION

III-A. Gender differences 90

III-B. Urine preparation 92

III-C. Urine storage 95

III-D. Bacterial contribution 101

CONCLUSION 105

LITERATURE CITED 107

CHAPTER IV – Variation of metabolites in normal human urine

OVERVIEW 112

INTRODUCTION 113

EXPERIMENTAL PROCEDURES

IV-A. Human control sample collection 115
IV-B. Human sample handling 116
IV-C. Guinea pig control sample collection 117
IV-D. Sample preparation 118
IV-E. NMR analysis 118
IV-F. NMR quantificaiton 119
IV-G. Statistical analysis 121

RESULTS

IV-A. Human demographics 122
IV-B. Metabolite identification 122
IV-C. Normalization to creatinine 123
IV-D. Inter-individual variability 124
IV-E. Gender 127
IV-F. Sampling time 127
IV-G. Intra-individual variability 128
IV-H. Guinea pig variability 131
DISCUSSION AND CONCLUSION 136
LITERATURE CITED 144

CHAPTER V – Metabolomic investigation of cellular biochemistry and pulmonary pathophysiology

OVERVIEW 148
INTRODUCTION 149

EXPERIMENTAL PROCEDURES

V-A. Volunteers 151

V-B. Isolation and purification of eosinophils and neutrophils 151

V-C. NMR spectral analysis of resting and stimulated cells 152

V-D. Sputum collection and processing 153

V-E. Cell counts 155

V-F. NMR spectroscopy of sputum 155

V-G. Statistical analysis 156

RESULTS

V-A. Chlorination of free tyrosine residues 157

V-B. Bromination of free tyrosine residues 159

V-C. Modified tyrosine residue detection by NMR in patient sputum samples 162

DISCUSSION 171

LITERATURE CITED 175

CHAPTER VI – Animal model of asthma

OVERVIEW 178

INTRODUCTION 179

EXPERIMENTAL PROCEDURES

VI-A. Guinea pig animals 182

VI-B. Guinea pig treatment 183

VI-C. Airway responsiveness to histamine injections 183
VI-D. Assessment of pulmonary inflammation 184
VI-E. Histological evaluation 185
VI-F. Histological identification of eosinophils 185
VI-G. Histological analysis 186
VI-H. Guinea pig urine collection 186
VI-I. Urine sample preparation 187
VI-J. NMR analysis 187
VI-K. NMR quantification 188
VI-L. Multivariate statistical analysis 188
VI-M. Statistical analysis 189

RESULTS

VI-A. Pulmonary function 190
VI-B. Ovalbumin sensitization and challenge induces airway hyperreactivity 190
VI-C. Ovalbumin sensitization and challenge induces airway inflammation 191
VI-D. Airway tissue inflammation 192
VI-E. Multivariate statistical analysis 193
VI-F. NMR and biochemical interpretation 194
DISCUSSION 194
LITERATURE CITED 197

CHAPTER VII – Analysis of acute and post-treatment human asthma

OVERVIEW 201
INTRODUCTION 202

EXPERIMENTAL PROCEDURES

VII-A. Acute asthma patient recruitment 203
VII-B. Sample preparation 204
VII-C. NMR analysis 204
VII-D. Exporting NMR spectra for analysis 205
VII-E. Multivariate statistical analysis 205
VII-F. NMR quantification 206

RESULTS

VII-A. Multivariate classification 206
DISCUSSION 211
LITERATURE CITED 216

CHAPTER VIII – Pediatric asthma longitudinal study

OVERVIEW 217
INTRODUCTION 217

EXPERIMENTAL PROCEDURES

VIII-A. Pediatric patient recruitment 219
VIII-B. Sample preparation 221

VIII-C. NMR analysis 221
VIII-D. Multivariate analysis 222
VIII-E. NMR quantification 222
VIII-F. Statistical analysis 223

RESULTS

VIII-A. Pulmonary function 223
VIII-B. NMR metabolite quantitation 224
VIII-C. Multivariate separation 226
DISCUSSION 226
LITERATURE CITED 229

CHAPTER IX – Conclusion

OVERVIEW 232
SUMMARY 232
DISCUSSION 236
LITERATURE CITED 244

LIST OF APPENDICES

Page

APPENDIX A – Multivariate analysis of NMR metabolomic data

OVERVIEW 246

INTRODUCTION 246

Appendix A.1 Data export 248

Appendix A.2 Binning bifurcation of resonance signatures 253

Appendix A.3 Resonance shifts 255

Appendix A.4 Principle components 255

Appendix A.5 Bin normalization 257

Appendix A.6 Data conversion 258

Appendix A.7 Multivariate statistical analysis 259

Appendix A.8 Statistics 260

APPENDIX B – Novel biomarker identification

OVERVIEW 262

INTRODUCTION 262

Appendix B.1 1D-NMR 263

Appendix B.2 Multidimensional NMR 267

Appendix B.3 Chemical standard 269

Appendix B.4 Double quantum filtered 1D-NMR 270

Appendix B.5 Mass spectrometry 271

Appendix B.6 Unique metabolite in guinea pig urine 271

CONCLUSION 274

APPENDIX C – NMR spectral baseline

OVERVIEW 275
INTRODUCTION 275
Appendix C.1 Classic NMR concentration determination 277
Appendix C.2 ^{13}Carbon 279
Appendix C.3 Peak overlap 281
Appendix C.4 Macromolecules 282
Appendix C.5 Baseline quantitation 285
Appendix C.6 Quantitation macro Lines 287
Appendix C.7 Biochemical definition of baseline 289
Appendix C.8 MS of urine macromolecules 290
CONCLUSION 291

APPENDIX D – Binding kinetics of calcium, cardiac troponin I peptide and cardiac troponin C

OVERVIEW 292
INTRODUCTION 293

EXPERIMENTAL PROCEDURES

D-A. Sample preparation 296
D-B. Ca^{2+} titration of ^{15}N-cTnC 298
D-C. cTnI$_{34-71}$ titration of ^{15}N-cTnC•3Ca^{2+} 298
D-D. cTnI$_{147-163}$ titration of ^{15}N-cTnC•3Ca^{2+}•cTnI$_{34-71}$ 299

D-E. cTnI$_{128-147}$ titration of ^{15}N-cTnC•3Ca^{2+} 299

D-F. NMR spectroscopy 300

RESULTS

D-A. Ca^{2+} titration of cTnC 300

D-B. cTnI$_{34-71}$ titration of cTnC•3Ca^{2+} 310

D-C. cTnI$_{147-163}$ titration of cTnC•3Ca^{2+}•cTnI$_{34-71}$ 315

D-D. cTnI$_{128-147}$ titration of cTnC•3Ca^{2+} 316

DISCUSSION 317

LITERATURE CITED 322

APPENDIX E – Database

OVERVIEW 327

INTRODUCTION 327

Appendix E.1 Patient information 329

Appendix E.2 Sample inventory 335

CONCLUSION 343

LIST OF TABLES

Page

CHAPTER II – Optimization of NMR qualitative and quantitative analysis of body fluids

Table II-I. Metabolite concentration determination via resonance integration and T_1 correction of Sample 2 65

Table II-2. Metabolite concentration determination via resonance integration with T_1 and filter correction 72

Table II-3. Concentration determination by Chenomx software 75

CHAPTER III – Urine as a biofluid for metabolomic studies

Table III-1. Urine metabolites 89

Table III-2. Female urine preparation 91

Table III-3. Male urine preparation 92

Table III-4. Female urine storage 96

CHAPTER IV – Variation of metabolites in normal human urine

Table IV-1. Urine metabolite concentrations 123

Table IV-2. Female urine metabolite variation over thirty days 129

Table IV-3. Guinea pig urine metabolite concentrations 136

CHAPTER V – Metabolomic investigation of cellular biochemistry and pulmonary pathophysiology

Table V-1. Patient characteristics, sputum cell counts, and spirometry 155

Table V-2. Bivariate correlation of CF patient oxidative metabolites 168

CHAPTER VI – Animal model of asthma

Table VI-1. Accuracy of guinea pig diagnosis by multivariate analysis 194

CHAPTER VII – Analysis of acute and post-treatment human asthma

Table VII-1. Multivariate diagnosis using aliphatic NMR signatures 208

Table VII-2. Multivariate diagnosis using aromatic NMR signatures 209

Table VII-3. Aliphatic spectral data points that separate populations 210

Table VII-4. Aromatic spectral data points that separate populations 210

LIST OF FIGURES

Page

CHAPTER II – Optimization of NMR qualitative and quantitative analysis of body fluids

Figure II-1. Schematic of five different 1D ^{1}H-pulse sequences 51

Figure II-2. 1D ^{1}H-NMR spectra acquired by different pulse sequences 57

Figure II-3. Differences in solvent suppression 58

Figure II-4. Longer acquisition time allows relaxation 61

Figure II-5. Lengthening delay time allows for fuller DSS relaxation 62

Figure II-6. T_1 relaxation of multiple metabolites in one sample 63

Figure II-7. Evidence of different T_1 relaxation in one metabolite 64

Figure II-8. Resonance attenuation based upon chemical shift 68

Figure II-9. Elliptical filter attenuation of signal intensity 70

Figure II-10. 'Brick wall' filter attenuation of signal intensity 71

CHAPTER III – Urine as a biofluid for metabolomic studies

Figure III-1. Metabolite changes in raw and filtered urine 94

Figure III-2. Female urine stored in a deep-freeze and at room temperature 98

Figure III-3. Spun female urine stored in a deep-freeze and at room temperature 100

CHAPTER IV – Variation of metabolites in normal human urine

Figure IV-1. Spectral analysis by Chenomx software 120

Figure IV-2. Creatinine variation in a normal human population 121

Figure IV-3. Human urine hippurate concentration normalization 125

Figure IV-4. Human urine citrate concentration normalization 126

Figure IV-5. Intra-individual urine citrate variability 130

Figure IV-6. Intra-individual urine tyrosine variability 131

Figure IV-7. Guinea pig urine creatinine variability 133

Figure IV-8. Urine hippurate normalization in control guinea pigs 134

Figure IV-9. Normalization of urine citrate in control guinea pigs 135

CHAPTER V – Metabolomic investigation of cellular biochemistry and pulmonary pathophysiology

Figure V-1. Formation of 3-chlorotyrosine by stimulated neutrophils 158

Figure V-2. Formation of brominated tyrosine in stimulated eosinophils 161

Figure V-3. Spectra of sputum from CF and control subjects 163

Figure V-4. Oxidized tyrosine formation in CF and normal subjects 166

Figure V-5. Correlations between cellular infiltration and 3-chlorotyrosine 170

CHAPTER VI – Animal model of asthma

Figure VI-1. Antigen challenged animals develop airway hyperresponsiveness 191

Figure VI-2. Antigen challenge induces inflammatory cell influx to airway lumen 192
Figure VI-3. Antigen challenge induces eosinophil influx into airway walls 193

CHAPTER VIII – Pediatric asthma longitudinal study

Figure VIII-1. FEV1 correlates with clinical presentation between visits 224
Figure VIII-2. Difference of 1-methylhistamine correlates with clinical follow-up 225
Figure VIII-3. 1-methylhistamine predicts follow-up presentation 226

APPENDIX A – Multivariate analysis of NMR metabolomic data

Figure A-1. Highlighted regions of stacked guinea pig spectra 249
Figure A-2. Macro for autoprocessing and exporting multiple spectra 251
Figure A-3. Macro for processing and exporting multiple spectra 252
Figure A-4. Stacked spectra with shifted highlighted region 254
Figure A-5. Spectral axis defined as ppm, Hz, and data points 259

APPENDIX B – Novel biomarker identification

Figure B-1. 1D ^{1}H NMR spectra of the unknown compound 264
Figure B-2. Schematic of the 1D-NOE pulse sequence 265
Figure B-3. 1D-NOE and 1D ^{1}H spectra of asthma urine 267

Figure B-4. 2D and 1D NMR spectra of the unknown metabolite in human urine 268

Figure B-5. 1D ^{1}H-NMR spectra of chemical standards 270

Figure B-6. Spectral analysis of a guinea pig urine NMR spectrum 272

Figure B-7. Chemical structure of the anesthetic urethane 273

APPENDIX C – NMR spectral baseline

Figure C-1. Quantitation of resonant peaks by integration 278

Figure C-2. Spectral richness and resonance overlap 279

Figure C-3. Spectral integration: effect of window size 280

Figure C-4. Resonance overlap in a 1D ^{1}H-NMR spectrum 281

Figure C-5. Sample filtration improves baseline 283

Figure C-6. Effect of urine filtration on NMR spectral lineshape 284

Figure C-7. Quantification by circumventing baseline 286

Figure C-8. Post-acquisition baseline removal by computer modeling 287

Figure C-9. Lines macro created for automated quantitation 288

Figure C-10. Pre-acquisition chemical removal of NMR baseline 290

APPENDIX D - Binding kinetics of calcium, cardiac troponin I peptide and cardiac troponin C

Figure D-1. Titration of human cTnC with Ca^{2+} 302

Figure D-2. One-dimensional traces demonstrating Ca^{2+} binding 304

Figure D-3. Calcium binding curves for the C- and N-domains of cTnC 305

Figure D-4. Computer simulated and empirical binding traces for Ca^{2+} and cTnC 309

Figure D-5. Titration of $cTnC{\bullet}3Ca^{2+}$ with cTnI peptides 311

Figure D-6. One-dimensional ^1H traces representing cTnI peptides binding to cTnC 313

Figure D-7. Computer simulated and empirical 1D ^1H-NMR traces for binding constants of cTnI peptides with cTnC 314

APPENDIX E - Database

Figure E-1. Principle investigator 330

Figure E-2. Patient clinical information 332

Figure E-3. Analytical textbox 333

Figure E-4. Sample storage page 335

Figure E-5. Sample inventory main page and sample summary 337

Figure E-6. Patient information 338

Figure E-7. Clinical information 339

Figure E-8. Diary information 340

Figure E-9. pH sample processing 341

Figure E-10. pH sample preparation 342

Figure E-11. pH sample acquisition 343

LIST OF ABBREVIATIONS

ANOVA – Analysis of variance

BAL – Bronchoalveolar lavage

CB – Cytochalasin B

CCDB – Cyber cell database

CF – Cystic fibrosis

CFTR – Cystic fibrosis transmembrane conductance regulator

COPD – Chronic obstructive pulmonary disease

COSY – Correlation spectroscopy

CPMG – Carr-Purcell-Meiboom-Gill pulse train

CSF – Cerebrospinal fluid

cTnC – Cardiac troponin C

cTnI – Cardiac troponin I

DISPI – Decoupling in the presence of scalar interactions

DSS - Disodium-2, 2-dimethyl (2-silapentane-5-sulphonate)

DQF-COSY – Double quantum filtered correlation spectroscopy

EDN – Eosinophil derived neurotoxin

Eos – Eosinophil

EPO – Eosinophil peroxidase

ER – Emergency room

EXV – External cross validation

PBS – Phosphate buffered saline

PCA – Principle component analysis

PLS – Partial least squares

Ppi – Pulmonary inflation pressure

FEV$_1$ – Forced expiratory volume in one second

FID – Free induction decay

fMLP – f-Met-Leu-Phe

FVC – Forced vital capacity

GC – Gas chromatography

GM-CSF – granulocyte/macrophage colony-stimulating factor

HMDB – Human metabolomic database

HOBr – Hypobromous acid

HOCl – Hypochlorous acid

HPLC – High performance liquid chromatography

HSQC – Heteronuclear single quantum coherence

ICS – Inhaled corticosteroids

IL – interleukin

K$_D$ – Dissociation rate

K$_{off}$ – Off-rate

LC – Liquid chromatography

LDA – Linear discriminate analysis

LOO – Leave-one-out

Lym – Lymphocyte

Mac – Macrophage

MBP – Major basic protein

MPO – Myeloperoxidase

MRD – Magnetic resonance diagnostics

MRI – Magnetic resonance imagine

MRS – Magnetic resonance spectroscopy

MS – Mass spectrometry

NANUC - Canadian National High Field NMR Centre

Neut – Neutrophil

NMR – Nuclear magnetic resonance spectroscopy

NOESY – Nuclear Overhauser effect spectroscopy

OVA – Ovalbumin

RDP – Relative distance plane

ROESY – Rotational NOE spectroscopy

SEM – Standard error of the mean

sTnC – Skeletal troponin C

sTnI – Skeletal troponin I

sw – Sweep width

T_1 – Longitudinal relaxation (spin lattice)

CHAPTER 1

Challenges in the application of ^{1}H-NMR spectroscopy to human metabolomic studies: lessons from asthma

I.A INTRODUCTION

The release of the human genome sequence, lead by the International Human Genome Consortium, occurred around the time I began my Ph.D. research [1](2001). The development and application of various techniques and technologies to tackle the mapping of the human genome, and the excitement of such a scientific endeavor, lead to an explosion in what is now termed the 'omics' era[2-6]. Genomics became the term to describe scientific investigations that map an organism's genetic material and follow the expression patterns of DNA[7-11]. The Human Genome Project made it possible to identify, track, and understand diseases that are associated with specific regions of human DNA. Proteomics took the science of large pathways and interactions a step further as it attempted to identify all of the information transcribed and translated from the DNA, now in the form of proteins[12, 13]. Proteomics is considered to be more inclusive in its biological information about an organism when compared to genomics because it includes not only the raw proteins translated from the DNA, but also all of the possible post-translational modifications (*e.g.* glycosylation, phosphorylation, protein cleavage, etc.)[13, 14]. Some scientists believe that proteomics, in comparison to genomics, advances the scientific investigation one step closer to an understanding of the physiology of an organism.

Recently, genomic and proteomic investigations have been followed by metabonomics. Along the same premise as proteomics, metabonomics attempts to take the biological investigation closer to the physiology or the phenotypic end-point of gene expression. Metabonomics attempts to track the flux of metabolites through the various anabolic and catabolic processes of an organism[15]. Typically, metabonomics monitors the biological response of a whole organism to a stimuli, stress, toxin, or disease[16]. A similar term, metabolomics, describes the study of metabolites and their continual rise and fall within a particular organ, tissue, or body fluid[17-19]. Metabolomics focuses the investigation from a 'whole animal' metabolic response to a more isolated system of study[20]. As the field of investigation matures the definitions for the terms metabonomics and metabolomics continues to be refined. For this thesis the term metabolomics will be primarily used. When I began my work in this field, an Internet search of publications on PubMed using the keywords metabonomics and metabolomics returned a total of 2 papers. As of July 2006, a similar search with the keywords metabonomics and metabolomics returned 146 and 333 papers, respectively.

Different groups of scientists have researched metabolic pathways for many years[21]. Biochemistry, nutrition, cell biology, medicine, and physiology are some of the key scientific disciplines that have devoted decades to understanding specific metabolic pathways. Each discipline often has a final scientific goal that relates to the area of specialty, and the study is often focused on a particular metabolite, enzyme, or cellular flux. The focus of metabonomic and metabolomic research is to gather and incorporate as much information as possible from multiple metabolic pathways to develop a more inclusive understanding of the biological system, tissue, or organism. This requires the involvement of multiple areas of science, the identification and discovery of many different metabolic pathways, and the identification and quantification of sometimes hundreds of metabolites and biomarkers. For human metabonomic studies the identified pathways and biomarkers can provide information beyond the endogenous biology, such as the influence of xenobiotics and pharmaceuticals[22, 23]. As well, the unique metabolites and metabolic flux may correlate with normal biological function[24], specific insult, stimuli, or disease[25, 26].

Metabol(n)omic studies have investigated metabolic flux and disease processes in single-celled organisms[27-29], plants[30], animals,[31, 32] tissue[33], organs[34], and human subjects[35, 36]. Studies have tried to identify pathways influenced by exposure to pathogens[35, 37], toxic chemicals[38, 39], pharmaceuticals[40, 41], and common variables such as diet[42]. They have also tried to identify useful biomarkers for diseases such as cardiovascular or digestive disorders[43]. The University of Alberta has supported the projects CyberCell (CCDB) and the Human Metabolomic Database (HMDB, www.hmdb.ca), which serve as examples of metabonomic and metabolomic studies, respectively. Project CyberCell is attempting to catalogue enough *in vivo* and *in vitro* information about the bacteria *Escherichia coli* so that computer programs can predict a cells response to a specific stimuli or stress. HMDB is currently cataloguing detailed information about small molecule metabolites found in different body fluids and will serve as a metabolic database for other laboratories. These two large-scale projects are examples of the broad (whole cell modeling) and focused (specific metabolites in a particular biofluid) studies identified as metabonomics and metabolomics, respectively. In a span of 5 years metabonomic and metabolomic research has grown considerably and with it the use of high-resolution nuclear magnetic resonance (NMR) spectroscopy to identify and follow metabolites of interest in a qualitative and quantitative manner.

NMR spectroscopy has been a central analytical technique in most chemistry laboratories since its invention over 60 years ago. The benefits of NMR analysis includes its non-destructive nature, the relatively short time required for data acquisition, minimal sample preparation, as well as a wealth of experiments that help to illuminate molecular structure and the dynamic nature of molecular interactions. The principle weakness of NMR is the limited sensitivity; 1D ^{1}H-NMR can only confidently detect and quantify metabolites greater than roughly 1μM. Technological advancements continue to increase the sensitivity and improve the overall stability of the spectrometer, ease of data collection, and increase resolution. These spectroscopic improvements have seen a concurrent widening of the applicability of NMR analysis. Currently, NMR is not only a principle technique for protein structure determination, but is used heavily for drug development, basic biological research, and metabolomics.

With an increase in the applicability of NMR technology came an increase in the accessibility for new scientific investigations. NMR is no longer found primarily in structural biochemistry or organic chemistry laboratories, but is expanding into physiological and clinical laboratories as well. NMR has become one of the primary tools for metabolic research. A similar development of technologies is evident in the growth of magnetic resonance imaging (MRI) and more specifically, magnetic resonance spectroscopy (MRS), which is used in metabolic research[44, 45]. These technologies allow for the identification of specific metabolites in the human body.

When I began my graduate degree I wanted to take the technology and theory behind NMR, which was taught to me from a thermodynamic and structural biochemistry background, to a more 'clinical' or 'biological' platform. An early collaboration was formed with Drs. Lopaschuk and Clanachan (Directors of Cardiovascular Research, and Pharmacology, respectively) to study cardiac metabolism pre- and post-ischemic insult and following pharmacological intervention. Another project with Drs. Lacy, Musat-Marcu, and Moqbel, from the Pulmonary Research Group, and John Bagu from the Canadian National High Field NMR Centre (NANUC), investigated pulmonary disease states through NMR analysis of sputum samples. Additional studies progressed as collaborations were formed with members of the University of Alberta; Drs. Marrie (Dean of Medicine), Adamko (pediatric pulmonologist), and Rowe (Director of Emergency

Medicine Research), for pulmonary studies of bronchiolitis, COPD, human normal, pneumonia, and asthma. The need to concentrate my research on a manageable project lead to a focus on the metabolomics of asthma.

Asthma is a pandemic disease with mortality rates that have increased significantly over the last 20 years[46, 47]. In Canada, asthma accounts for approximately 500 deaths annually and hospitalization rates have quadrupled over the last few decades[48]. The pathophysiological characteristics of asthma are heterogeneous, which has lead to difficulties in clinical diagnosis and treatment, as well as complicating basic research. It was hypothesized that through a metabolomics approach high-resolution NMR analysis of biofluids might identify key metabolites that are indicative of various phenotypes of asthma, as well as enhance the current understanding of asthma and response to treatment. A more detailed background of asthma, as well as some of my investigations into the pathophysiology of the disease will be provided in Chapters V to VIII.

To apply NMR spectroscopy to human metabolomic studies, many questions needed to be addressed regarding the applicability of NMR, how to address the disease being studied, the method for sample collection and preparation, the metabolic behaviour of normal subjects, the extraction of quantitative and qualitative information from NMR spectra, as well as the proper statistical methods required to determine biochemical significance. In addition, a number of philosophical issues needed to be considered; what is ‘human normal’, 'toward what end are metabolomic studies – simple population separation/identification, or a detailed biochemical understanding of a metabolic or disease state'? This thesis will outline the technical and philosophical challenges I encountered when applying high-resolution ^{1}H-NMR spectroscopy to human metabolomic investigations.

I.B METABOLOMIC TECHNOLOGIES

Several methods have demonstrated their usefulness in the identification and quantitation of various compounds for metabolic research. The most common techniques include mass spectrometry (MS)[49-51], high performance liquid chromatography (HPLC)[52, 53], gas chromatography (GC)[54-56], and NMR spectroscopy[33, 57-59]. Each of these technologies has inherent advantages and disadvantages. Of the more common technologies used for metabolite detection, MS and NMR differ significantly from one another and when the strengths of each are utilized appropriately, compliment each other towards a fuller understanding of a metabolic profile.

I.B.1 Mass Spectrometry

The primary strength of MS is the capability to detect metabolites in femtomolar concentrations. However, MS has limitations that impair its use for many metabolomic investigations. MS is unable to differentiate isomers, involves extensive sample preparation prior to analysis, metabolites often need to be derivatized to aid in detection, but often can not be quantified, and the sample is destroyed in the process. These limitations significantly restrict the applicability of MS to most metabolomic investigations.

I.B.2 Nuclear Magnetic Resonance Spectroscopy

High-resolution ^{1}H-NMR spectroscopy has a number of advantages that lend itself to the successful application to metabolomic studies. NMR is non-destructive, meaning that the liquid or tissue samples may be returned to long-term storage following data acquisition. This is extremely important, especially during preliminary studies, because the same samples may need to be returned to the spectrometer at a later date if techniques or technologies improve. For most biofluids in metabolomic investigations there is little sample preparation required prior to NMR data acquisition (no need for separation, purification, or metabolite derivatization). In addition, NMR has a wide range of data acquisition parameters, which allows scientists to identify compounds based upon unique characteristics, such as diffusion, or additional information that may be acquired through 2D NMR. The data, often a 1D ^{1}H-NMR spectrum, is both qualitative and quantitative. The NMR spectrum allows for the separation of metabolites according to their unique spectral signatures (resonant frequencies and scalar couplings) leading to the qualitative identification of, in the case of urine, hundreds of metabolites in a single spectrum. To extract quantitative information from 1D ^{1}H-NMR the area under the respective resonant peaks of a metabolite may be integrated along with the resonance of a standard of a known concentration. The non-destructive qualitative and quantitative analysis of metabolites found in biofluids is the key driving force behind the increased use of NMR for metabolomic studies.

One of the issues that needed to be addressed with metabolomic investigations was the increase in the number of samples to be analyzed by NMR. In contrast to traditional protein NMR, samples are not purified proteins undergoing structure determination over a period of several months. Metabolomics requires the analysis of hundreds of samples from human or animal subjects with similar pathophysiologies. Data acquisition must provide optimal information to allow for accurate qualitative and quantitative analysis. Metabolomic investigators require pulse sequences that provide flat spectral baselines, optimum solvent suppression, superior signal-to-noise, and improved resolution while remaining robust enough to analyze hundreds of samples without requiring lengthy pre-acquisition and pulse sequence set-up[60, 61].

Since metabolomics is the study of biological processes, often in humans, the samples are extremely important, difficult to obtain, and the acquisition and interpretation of the information is vital. Therefore, the sampling of the particular biological pathway has been the source of a lot of my work and many discussions.

I.C METABOLOMIC SAMPLES

Metabolomic studies involve monitoring biochemical and metabolic pathways; therefore requiring sampling of biological processes in the organism. When sampling a specific pathway or disease state in an organism there are a number of considerations to be addressed.

I.C.1 Sampled Population

The first challenge is the selection of appropriate populations to be included in the study. My work has included the study of a number of different diseases and biological systems. I have looked at pulmonary disorders such as tuberculosis, viral and bacterial pneumonia, cystic fibrosis, chronic obstructive pulmonary disease, bronchiolitis, and asthma. With each disease it is important to understand the subjects that will comprise the test population. Researchers must address different factors that relate to the particular pathophysiology that will ultimately dictate the inclusion and exclusion criteria for a sampled population. As an example, asthma is a heterogeneous disease and researchers must be rigorous to define the disease characteristics, patient symptoms, and 'type' of asthma they wish to study (*i.e.* atopic, steroid resistant, pediatric vs. adult, outpatient vs. acute etc.). The choice of disease phenotype and study greatly impacts the sample population and sampling methods.

I.C.2 Biological Relevance

Researchers must decide upon the most appropriate and acceptable method of sampling. With a focus on pulmonary patients there are different possibilities for sample collection and biofluids analyzed. Initially, I sampled the human lung through the collection of sputum[62]. Later, I focused on human urine as the biofluid for NMR analysis[63]. For pulmonary investigations, these two biofluids, sputum and urine, possess very unique properties and represent the quintessential interplay between key issues of human sample collection; invasiveness, patient cooperation, and biological relevance.

Often for metabolomic studies each investigation attempts to elucidate a shift along a particular biochemical/metabolic pathway or the clinical manifestation thereof. The researcher must ensure that the biological sample is appropriately captured and reflects the pathway, or disease in question. To aggravate this question further, direct sampling of the pathways is often not an option. The direct sampling of a diseased tissue may not be realistic due to the isolation of the event to a particular organ that is inaccessible to the majority of sampling techniques. Direct sampling may also be difficult if the site of the disease or metabolic reaction is essential to the test animal or human subject and therefore may not be removed or even sampled. Therefore, biological samples that are relevant to the pathways being studied are not always biologically feasible.

I.C.3 Sampling Invasiveness

Similarly, another factor to be considered, especially in human studies, is the invasiveness of biological sampling. This may be bypassed with animal models where there is greater methodological freedom, as an example; with an animal model it becomes possible to harvest whole organs. However, in animal models investigators take a step away from human relevance and may compromise the correlation to human disease for easier access to biological tissue/fluid samples. Often in preliminary studies invasiveness of biological sampling is essential in order to prove relevance. Some studies sample directly from the site of interest in human subjects through tissue biopsy. In some of my initial work I used sputum collection. This direct sampling allowed for identification of relevant pathways and the correlation of key indicators with the processes of the upper airways. Later studies attempted to identify remnants, or secondary-indicators of biological events from more accessible biofluids, such as urine; sites that allow for less invasive methods of sampling are often more acceptable for human studies.

For example, in my first study I sampled induced sputum in patients with cystic fibrosis (CF) and asthma. Although sputum collection samples the ‘lung’, with minimal contamination from the upper respiratory tract and oral cavity, the technique is rather time consuming, and can be a source of irritation or medical risk for patients in respiratory distress. In addition, collection of sputum from control human subjects was not possible due to non-productive collection sessions. Although sputum collection provided a direct sampling method of the airways, and perhaps the most biologically relevant, the limited number of subjects that produced a sputum specimen dictated that another sampling method was needed.

The difficulties of biological relevance and invasiveness of sampling procedures have plagued medical investigations for sometime and in some instances technological advancements have allowed for their circumvention[43, 64]. For example, MRI has allowed the visualization of the whole human body, without biofluid or tissue sampling, including key organs that are not readably accessible by traditional techniques. MRI allows for the visualization of tissue and organs such as the human brain[65], blood vessels[66], internal organs[67], and even the detection of selected metabolites (MRS)[68]. However, there are some limitations to MRI; often contrast agents are required, quantitation is not possible, or at best is a rough estimate, and some organs are not visible, in particular the lung[69]. When direct sampling, visualization, or detection of a biological process is not possible researchers must employ more creative techniques to monitor pathways of interest.

As mentioned before, the population to be sampled can limit accessibility because of the degree of invasiveness. Ethics concerning human investigations limits the degree of stress researchers may place on the subject being sampled. As well, when studying a particular pathophysiology the disease itself will impact the availability of various biofluids.

I.C.4 Ethics

The issues of human research ethics and subject compliance are related to the benefits of direct sampling for biological relevance and the limitations of invasiveness. Human studies must conform to established guidelines of research protocols and professional codes of conduct. The appropriate treatment of human subjects, as well as the use of samples and information in basic and clinical research studies must ensure the safety, integrity, and privacy of the human subject. There are additional restrictions placed upon researchers regarding their methods of sampling, analysis, and the release of final conclusions.

I.C.5 Subject Compliance

Investigators must also consider subject compliance. Compliance, or study completion, can be a challenge, especially for investigations of human diseases where the patients undergo a changing treatment regimen. Therefore, human studies must be as short as possible and involve a sampling method that is acceptable to the individual. It is the choice of the individual if they wish to participate in the study; this consideration can limit the design of metabolomic and other clinical investigations.

I.C.6 Sample Type

Finally, the biological sample must be appropriately prepared for analysis. In Chapter III I will discuss sample preparation, handling, and different biofluids in greater detail. Investigators must identify inherent physical limitations of the sample and whether or not metabolic analysis will be possible. My expertise in sample analysis is with NMR, specifically high-resolution liquid ^{1}H-NMR. This focus on NMR limits the types of samples I may analyze, the compounds or metabolites I can identify, and the scientific questions I can answer. Liquid NMR has been the principle sampling method for decades; however, solid-state NMR is also a choice for analysis and has been used in some biological studies (*e.g.* tissue samples taken from cancer patients). The advantages of liquid and solid-state NMR must be considered for each investigation; in my work liquid state NMR has been the focus.

Some of my initial pulmonary studies used sputum as the biofluid analyzed by liquids NMR. The physical nature of sputum presented significant challenges during spectral acquisition, and later qualitative and quantitative analysis. Due to the semi-solid and highly viscous nature of sputum additional time was required for sample preparation, metabolite identification, and concentration determination. A detailed discussion of sputum analysis will occur in Chapter V.

1.C.7 Sample Preparation and Storage

When the above considerations have been addressed the researcher must prepare and store the sample prior to, or following analysis. This is a significant matter, particularly with human studies in a clinical setting. Often samples are collected during a patients' visit to the hospital and must be stored on site for a certain amount of time. The samples are then transferred to a central storage facility to await analysis and final long-term storage. Each step in sample preparation, transport, and storage has possible effects upon sample integrity, biochemical identification, and correlations with disease states. Chapter III demonstrates the extent sample preparation and storage have on metabolites in human urine and highlights the need to carefully manage sample integrity throughout a study[63].

I.D URINE: A METABOLOMIC BIOFLUID

Overall, many different biofluids have been used for metabolomic studies. Bronchoalveolar lavage (BAL)[70-72], sputum[62, 73, 74], cerebrospinal fluid (CSF)[17, 75], breath condensate[76, 77], blood[78], serum[79, 80], and urine[81, 82] are a few of the more widely studied samples taken from animal and human subjects. As mentioned before, the biological system under investigation must be considered when choosing the appropriate biofluid. For my pulmonary studies there were concerns about invasiveness (sputum) and sample collection (lack of normal subject sputum) when sampling the human lung. Although urine does not sample directly from a patients' lung, it does provides a rich metabolite profile that behaves very well with the present techniques and technologies of liquid NMR (see Chapter II and IV). After contemplation of the advantages and disadvantages of urine, and the key considerations listed below, my later studies chose urine as the biofluid to be sampled from pulmonary patients.

1.D.1 Sampled Population

For human pulmonary investigations urine is not often thought of as a biologically relevant choice as a biofluid. However, with my focus on pulmonary patients presenting with many different pathologies urine was considered to be the most versatile and well-suited sampling method. Direct pulmonary sampling such as BAL, or sputum is ideal for metabolomic investigations of lung disorders; however, the invasiveness of the technique, the dangers of aggravating already ill patients, and the limited patient compliance removed these techniques as options.

The patient population for each of the pulmonary disorders studied had a significant impact on type of biofluid sampled. Many of the different pulmonary diseases present in older or weaker populations (*e.g.* pneumonia and COPD), which are unable to manage or recover from some of the more invasive sampling methods. In addition, my research had a significant pediatric asthma component. When considering the patients that would be recruited for the pulmonary studies urine was a more feasible option than other biofluids.

I.D.2 Biological Relevance

A concern that arose when sampling a patients' urine was the validity with regard to the pulmonary disorder. Are processes in the lung represented in the urine, or so eloquently put by my supervisor 'Is the lung connected to the urine'? Discussions arose concerning more specific experiments to assess the transition period of a number of different inhaled compounds, or their metabolites, to the later arrival in the urine. Unfortunately, time did not permit me to complete these studies. However, a number of investigations have looked at metabolites that originate in the lung and their subsequent appearance in the urine (*e.g.* human exposure to pollutants[83], inhaled radiotracers [84], and some biomarkers of asthma[85, 86]).

Another asset of urine is the wealth of metabolites that appear in the biofluid. The 1D ^{1}H-NMR spectral profile of urine is extremely complex with numerous resonant peaks throughout. Rough estimates currently list the number of peaks to several thousand, representing anywhere from 300 – 1000 different compounds in concentrations greater than 1μM. With regard to biological relevance, it is hypothesized that some metabolites and perhaps some currently unknown compounds may have biological relevance with ongoing pulmonary pathophysiologies.

I.D.3 Sample Acidity

A difficulty that arose in my research was the large pH variability observed in urine samples. The chemical shifts (resonant frequency) of metabolites identified by NMR are subject to perturbations as a result of changes in pH. The range of physiological urine pH values often rested between pH 6 to 8. For all NMR investigations in this thesis urine samples were pH'd to 6.8 with HCl and NaOH (see the *Methods* section of each Chapter). This standardization allowed for more consistent metabolite resonance identification and translation between multiple spectra. In addition, some spectral analysis software used for metabolite quantitation have multiple metabolite databases at different pH (Chenomx, Edmonton, AB, and ACD labs, Toronto, ON). This allowed for accurate metabolite identification and the calculation of concentrations at physiologically relevant pH intervals (*i.e.* pH 5 – 8).

I.D.4 Sample Preparation and Storage

It is important to remember the physiology from which the sample originates, as well as the method of collection. Urine is not a sterile fluid. Within urine there are a number of compounds and sources of energy with which microorganisms flourish[87, 88]. Many studies have looked at the propensity of infection in patients with intraurethral catheters[89-91]. It is common for individuals who are immunocompromised to develop urinary tract infections due to the ability for micro-organisms to flourish from the nutrients in urine[88, 91, 92]. As urine is excreted from the body it comes in contact with surfaces that are host to a number of *normal flora[93, 94]*. These bacteria can contaminate urine samples and therefore modify the metabolites before final analysis. Pulmonary patients in my studies visited the hospitals in either the Emergency Department or the Outpatient Clinic at the Universities of Alberta and Manitoba. During this time the patients provided a sample and proceeded with their appropriate clinical procedures and therapies. A risk of this method of sample collection was the possibility that the urine sample sat on a countertop for a period of time before it was frozen. In addition, there was a high probability that the urine would remain in the clinic freezer for a few hours before being taken to a laboratory deep-freeze (-80°C) to await final analysis. Depending on spectrometer usage, acquisition of the NMR spectra can be delayed by weeks. Each of these steps has the potential to cause further sample deterioration or modification to take place. It is not known to what extent samples degrade, change, or if bacterial contamination alters the metabolite profile before the sample is analyzed. I designed an investigation, which helps to outline appropriate urine sample collection, preparation, and storage techniques to ensure the final analysis represents the original metabolic profile of the subject (Chapter III)[63].

I.E HUMAN BIOLOGICAL HOMEOSTASIS vs. VARIATION

I.E.1 Biological Complexity

The human body is composed of a vast array of metabolic processes. It is astounding to consider the constant flux of metabolites and chemicals throughout the human body, through various vessels and organs, and within the organelles of every cell. The complexity of the organization is remarkable when one considers that the human body has approximately 10 – 100 trillion cells, with all of the biochemical processes of every cell controlled in some fashion. Basic biology teaches that one of the basic mechanisms of control is through the use of feedback loops. This process allows cells and the human body to coordinate the transport, conversion, and utilization of metabolites in a controlled and relatively efficient manner. This leads to the understanding of biological homeostasis. Homeostasis teaches that individuals cannot survive if all biological mechanisms of the human body are allowed to proceed un-checked. In order to maintain a cellular balance, and in order to perform specific tasks essential for life each of the metabolic pathways must remain under some sort of control.

1.E.2 Biological Homeostasis

We can easily observe examples of biological homeostasis, such as the regulation of body temperature. Human bodies maintain a constant temperature of 37°C (98.6°F). There are a number of mechanisms through which the human body regulates and maintains this temperature, including automatically closing pores (goosebumps) and rapidly contracting muscles (shiver) to increase body temperature, or perspiration to lower temperature. Too high (fever, 38°C), or low (hypothermia, 36.1°C) of a body temperature causes cellular and organ function to fail and can lead to death. Similar biological and homeostatic control is required at the biochemical level; for example, the regulation of blood glucose concentrations. If control over blood glucose concentration is lost an individual may develop diabetes, which if left untreated will lead to death. Therefore, it is clear that the human body is not simply a container in which metabolism is allowed to proceed randomly, but rather is a tightly regulated dynamic process operating within defined extremes. It was believed that since all biochemical processes must be maintained and regulated to create life and since all 'healthy' individuals are relatively phenotypically similar, *i.e.* must maintain similar basic biochemical functions, then there must be a basic biochemical definition of metabolism. As well, because we can already clinically define 'sick' from 'healthy', this basal metabolic definition must be known.

Surprisingly, little research has been conducted to define the normal and healthy excretion of metabolites in human urine. Often the assumption is that science already understands the excretion of metabolites from the human body, or there are citations to the Geigy tables, which try to define normal limits of physiological functions[95]. Careful review of scientific literature and textbooks reveals that very little is known regarding urine secretion for the majority of metabolites. Physiologically and clinically the attention has been given to a handful of ions and metabolites (*e.g.* urea, glucose, albumin, creatine/creatinine, Ca^{2+}, Mg^{2+}, and K^{+}). The often-cited Geigy tables reference original scientific articles that use only a handful of human subjects[95]. A recent project has begun at the University of Alberta called the Human Metabolite Database. This project will attempt to identify and set normal ranges on a number of metabolite concentrations found in various biofluids (www.hmdb.ca). Through my research it became clear that a better understanding of normal metabolite excretion was needed before any comparisons were made to a disease population.

1.E.3 Metabolic Homeostasis

A definition or a philosophical choice must also be made concerning what it means to be 'normal'. If asked, the majority of individuals would self-diagnose themselves as normal. If people are normal, they are physically stable, healthy, and therefore should have similar urine metabolite secretion. Clinically and intuitively it seems rather basic that a general healthy phenotype would produce similar urine metabolite profiles since at a basic level every healthy individual must perform similar biochemical reactions (*e.g.* glucose, protein, and nucleic acid metabolism). The phenotypic endpoint of many biochemical pathways, when altered, may be obvious in a medical context. Clinically, there is no need to provide a urine sample to differentiate a healthy individual from one undergoing an acute exacerbation of asthma. However, in my research there remained a need to define the metabolic profile of healthy human subjects. With the earlier assumption of biological homeostasis the biochemical definition of a normal population became very challenging. Under the gross physiology and phenotype of the healthy human population, is there an underlying metabolic and biochemical variability or flexibility that we are unaware of?

I.F NORMAL METABOLOMIC POPULATIONS

I.F.1 Importance of Normal

Since the historical periods of the Renaissance and The Enlightenment the Scientific Method has been the driving principle behind all scientific investigations. Experimental procedures attempt to isolate the system being studied as much as possible, measure and observe outcomes, and draw conclusions based upon the comparison of test measurements with those from a control or standard. Up to this point much of the focus has rested primarily upon the test population and the disease being studied. Often less attention has been awarded to the sampled population upon which all measurements are finally compared to and conclusions are based upon: the control or normal population. As I proceeded through my PhD research it became clear along that I needed to clarify, define, and/or remove as many assumptions as possible. One of the largest assumptions I came across was the idea that it was already known how a normal human population appears biochemically when sampled via the urine.

Defining the sampled population is particularly important in human studies where disease variability is compounded by genetic heterogeneity, variation in treatment, response to treatment, and socio-economic/lifestyle differences. Efforts were made to sample a 'similar disease population' so that measured observations may be correlated to the disease without significant deviations due to outside influences. The initial chapters of this thesis devote significant time to the consideration of analysis; how to appropriately sample patients and human subjects (Chapter III), and how to acquire and interpret the data (Chapter II) so that valuable information about the disease, physiology, or pathology may be extracted (Chapters V – VIII).

In some avenues of biological investigation the idea of a normal population might be easier to satisfy. As an example, in measuring cellular or a biological response to certain stimuli often the same animal or a cell line may be used in the control and test group. In other experiments it is possible to preserve the same subject throughout the experiment, *i.e.* take a baseline reading of a specific process, treat or stress the subject, and then repeat the measurement. These types of experiments are able to remove many of the questions regarding inter-subject comparison of control and test groups by maintaining the same subject with a basic "before and after" type of analysis. However, in human studies this is often not possible, particularly when trying to elucidate the altered biochemistry or physiology of a disease state.

I.F.2 Definition of Normal

When trying to identify a normal population for investigation we must question our own inclusion and exclusion criteria and carefully scrutinize why they were selected. There are a number of unique stressors that influence a sampled normal population. My initial 'normals' study took place at a large academic institution and medical facility where there were predisposed socio-economic levels unique to the geographic region. With the rise of immigration in the province of Alberta there has also been an introduction of individuals from different cultures and genetic backgrounds. As well, in a Western culture there are unique dietary trends and a larger reliance on over-the-counter pharmaceuticals (*e.g.* ibuprofen, acetaminophen, multivitamins). As these issues were considered a list of confounding factors grew that could potentially influence the metabolite concentrations observed in the sampled normal population.

To investigate the metabolites in urine from a normal adult human population we selected 30 male and 30 female subjects. They were sampled morning and afternoon of every day, for 30 days. The primary exclusion criterion for this study was an absence of major chronic illness. I chose to ignore all of the possible influencing factors mentioned in the paragraph above for the preliminary sampling of a normal population. Over the 30 day study some of the participants suffered a "cold" for a few days, others took vitamins, were social drinkers, took birth control, anti-depressants, medications etc. Each participant was required to keep a diary throughout the 30 days recording diet, supplements, activities, menstrual cycles, and stress levels. This detailed meta-information will allow for the additional exclusion of some individuals and re-analysis at a later time, but for the first study it was decided to focus on a more inclusive normal population (see Chapter IV).

Philosophically I believe the study described in Chapter IV is representative of normal. Some critics' contest that all participants taking pharmaceuticals should be removed. However, do we then remove anyone who had a "cold" for a period of time over the 30 days of the study? I argue that in our society it is normal to take the occasional pharmaceuticals for various ailments (*e.g.* ibuprofen, Sudafed®, etc), not participate in regular exercise, consume the occasional alcoholic drink, and experience high levels of stress. My study attempted to sample and study metabolite excretion in urine from a normal population, not some conceptualized ideal population. Inevitably, I would like to extend my metabolomic research to asthma urine sampled from patients visiting the ER or the outpatient clinic. As such it is of no use to determine metabolite excretion in an idealized healthy population, but rather an honest representation of the daily activities and lifestyles of regular human subjects that better incorporate the activities of a future test population (*e.g.* asthma patients).

I.G BIOMARKER DISCOVERY AND METABOLOMIC CORRELATIONS

Once a normal population has been carefully defined researchers can attempt to identify unique metabolic pathways that are altered in some way in a treated or diseased population, and correlate the observed changes to a particular altered state of physiology. Accurate identification and quantitation of key biomarkers are at the root of all metabolomic investigations. It was hypothesized that by identifying either unique metabolites, or by profiling a collection of metabolic pathways, insight may be gained regarding particular biological events associated with diseases that may lead to improvements in understanding, diagnosis, prognosis, and therapy.

I.G.1 Systems Biology, Complex Questions

"The failure of reductionism doesn't mark the failure of science, but only the replacement of an ultimately unworkable set of assumptions by more appropriate styles of explanation that study complexity at its own level and respects influences, …." Stephen Jay Gould[96]. Metabolomic investigations sampled at the biofluid and cellular level, acquire data at the atomic and biochemical level, and attempt to correlate the metabolite profile to the unique pathophysiology of the disease under investigation. This is an extremely complex process requiring knowledge of many different scientific disciplines. When does a metabolite cease being a simple molecule that is broken down or utilized in a biochemical manner and become a metabolite or biomarker important to clinicians and researchers? Beneath the simple idea of identifying metabolites in a metabolomic investigation is the desire to draw definitive scientific conclusions regarding complex physiology and disease biochemistry. Therefore, what information is required for a metabolomic study? A black-box multivariate statistical analysis that simply returns a statistically significant separation of two populations, or an in-depth biochemical interpretation of the data involving an understanding of analytical chemistry, biochemistry, physiology, microbiology, immunology, histology, medicine, statistics, and systems biology?

I.G.2 Metabolite or Marker?

A metabolite may represent a step along any portion of an extensive pathway of metabolic reactions in the human body. The unique biomarker or metabolite may either represent a primary causation or a secondary indicator of a particular pathway. The difficulty resides in the differentiation between a common metabolite, which represents basal biological processes, and a general indicator of 'biologic stress', or a biomarker, which represents a unique compound providing specific knowledge regarding the pathophysiology being investigated.

Often among the collaborations between basic researchers and clinicians there remains a discrepancy in the philosophical question regarding the final endpoint for metabolomic investigations. Is the purpose to separate populations based upon unknown 'peaks' and a linear combination of spectral signatures, or is the final purpose to address the metabolomic question at the biochemical level? Unlike the questions posed by clinicians, for the basic researcher a metabolomic investigation may continue past the identification of key metabolites to an understanding of the metabolic pathway and the interpretation of the biochemical information back to the pathophysiology of the disease being studied. This thesis attempts to address some early assumptions in the metabolomic field and optimizes scientific methods so that metabolomic analysis and interpretation can be drawn to a completed conclusion, allowing for the best understanding of the biological system under investigation.

1.G.3 Population Separation

Earlier in this chapter I introduced and provided examples of appropriate methods to correctly sample a biological system, analyze the biofluid by accepted analytical techniques, and identify a signature that is unique and provides relevant information when compared to a correctly formulated normal or control population. A pivotal question that I believe needed to be answered next was whether or not additional efforts needed to be taken once a unique spectral signature or metabolite was identified. Metabolomics has been heralded as a scientific investigation, which can directly sample biological endpoints (metabolites) that are the result of altered physiology and thus the disease in question. In a medical environment any clinical differentiation between two populations, healthy and diseased, is sufficient to warrant clinical intervention and the commencement of appropriate therapeutic measures. A number of statistical techniques have been developed that attempt to take amassed data and either blindly sift for unique signatures and statistical correlations, or through a neural network fashioned 'learn' from a training set of known diseases and apply those metabolic limits and trends to any future samples. A consideration that needs to be applied to metabolomic studies, and specifically to its clinical application, is simply can "two populations can be separated?".

Statistical analysis tools such as principle component analysis (PCA) [97, 98]and relative distance plane (RDP) [99-101]allow for the decomposition and projection of complex, multivariate, and high-dimensional data onto a simplified set of vectors that can be easily visualized and interpreted by researchers. Often this method of analysis requires little to no prior knowledge and relies upon machine learning to separate distinct populations. This type of analysis is very attractive to non-hypothesis driven research as it allows for the possible discovery of novel compounds. A more detailed discussion of multivariate analysis for NMR-metabolomic investigations is provided in Appendix C.

The method of multivariate statistical analysis, or data mining requires large datasets (patients) in order to reduce the dimensionality and refine the statistical conclusions to spectral signatures that separate the two populations. Without appropriately sized datasets multivariate analysis is prone to 'over-fitting', which limits the effectiveness of population separation when additional samples are studied. Many metabolomic studies, and most with a clinical interest, publish at this stage of investigation. The predominating question is whether the two populations are different and can be separated statistically. Often, the unique NMR signatures, that the PCA or RDP algorithms used to separate the populations, are not translated to actual metabolites. For many journal publications the metabolite concentrations are not reported, and the authors provide no definition of concentration ranges, or any explanation regarding specific pathways.

I.G.4 Biochemical Interpretation

The strength of a metabolomic study is the direct sampling of a human subject, the measurement and interpretation of final phenotypic outcomes of metabolic pathways, and the ability to trace those disease manifestations to a particular biochemical process. NMR spectra can be analyzed by specific metabolite identification, which requires prior knowledge of resonant frequencies and spectral signatures, or through multivariate statistical analysis. There is a methodological strength in combining the two techniques of analyzing NMR data acquired for metabolomic studies. Since my studies have primarily focused on human pulmonary disorders there is a preexisting wealth of knowledge that allows for a reasonable hypothesis regarding metabolic pathways unique to pulmonary disorders. For instance, there are known pathways, cellular responses, and immunological cascades that predominate in certain pulmonary pathophysiologies that lend themselves to metabolomic studies. Investigators are limited to known compounds, but may quickly and confidently qualitatively, and quantitatively analyze 1D ^{1}H-NMR spectra. Researchers should keep in mind that the prior knowledge is not complete and new compounds may be identified through a metabolomic investigation. The two methods of analyzing NMR generated data that allows for a more complete extraction of pertinent information. Appendix D will continue the discussion regarding the interpretation and the continuing efforts to translate 'spectral signatures' of interest as determined by multivariate analysis to biochemical compounds with known structure and physiological involvement.

In human urine the number of compounds that may be identified by 1D ^{1}H-NMR is theoretically over a thousand (including influences such as diet, pharmaceuticals, xenobiotics, and normal physiology). In any human urine sample the number of compounds present in a single sample, that may be easily identified and quantified by 1D ^{1}H-NMR, is estimated to be around 400. When identifying metabolites that may differentiate populations and diseases it is important to ensure that the compound is appropriate and unique to the disease, long-lived or stable, and is specific to a relevant pathway of the disorder.

I.H THESIS RATIONALE

Metabolomics is quickly becoming a popular method for the sampling of complex pathways in organisms. Metabolomic investigations attempt to identify metabolites that are either causative or indicative of unique biological processes or disease pathophysiologies. Therefore, these studies are extremely useful as a basic research attempt to enhance the current state of knowledge of a particular biological process. As well, because metabolomics often samples directly from human patients clinicians are interested in the application of metabolomic-generated data towards patient diagnosis, prognosis, and the managing of appropriate therapeutics. This thesis attempts to carefully develop methods of collection, analysis, sample preparation and storage, as well as statistical correlation and interpretation to ensure proper scientific investigation in metabolomic studies.

Specific Aim #1 – Optimization of NMR data acquisition for metabolomic investigations

Chapter II – NMR has become a leading analytical tool for metabolomic studies. There are significant differences in scientific methodology and acquisition requirements for metabolic studies that must be addressed. *This chapter reviews operating procedures for data acquisition, post-acquisition processing, and quantitative analysis.* Acquisition parameters are optimized to ensure the excellent quality of the data, and to minimize the influence of spectrometer engineering. Different software packages for metabolite quantitation are identified and compared for accuracy and precision.

Specific Aim #2 – Urine sample preparation and storage

Chapter III – Urine is a popular biofluid for human metabolomic studies because of the ease of collection and rapid time for analysis. It is necessary to identify that in human clinical investigations urine samples are often collected at hospitals where sample storage, preparation, and analysis can be delayed. *This chapter investigates the impact of sample preparation and storage techniques on the fidelity of the urine to accurately represent the original metabolic state of the human subject.*

Specific Aim #3 – Normal population

Chapter IV – The majority of clinical and metabolomic investigations attempt to identify changes in concentration and elucidate fluctuations in particular metabolic pathways that diagnose and provide information regarding a disease state. Data from the disease population is compared to a sampled population of 'normals' to identify changes in metabolite concentrations. *This chapter investigates metabolite variability in a sampled normal population over a period of 30 days ((inter-individual variability and intra-individual variability).*

Specific Aim #4 – Metabolomic analysis of sputum from CF patients

Chapter V – Traditional pulmonary diagnosis and assessment techniques for asthma or CF involve invasive procedures such as bronchoscopy, measures of overall lung function, and general immunological assessments such as sputum cell counts. *This chapter demonstrates the ability of NMR to analyze induced sputum from CF patients and identifies markers of immunological cellular activation.* This preliminary paper points to the beneficial information obtained by identifying cellular activity rather than simply cell presence in conjunction with a specific disease pathophysiology.

Specific Aim #5 – Guinea pig model of asthma

Chapter VI – The animal model of asthma allows for the control of biological history, environment, and therapeutics, something not possible with human clinical studies. Urine was collected from control, sensitized, and challenged guinea pigs. *This chapter presents the correlation of metabolic signatures from NMR spectra of guinea pig urine as an animal model of allergy and asthma.* The ability to separate differently treated animals by NMR analysis of the urine confirms the metabolic indices of underlying pulmonary physiology.

Specific Aim #6 – Analysis of acute and post-treatment human asthma

Chapter VII – The first application of metabolomic techniques towards the study of asthma through the analysis of human urine focused on patients admitted to the University of Alberta emergency department. The patients were sampled a second time following clinical treatment. *This chapter presents a metabolomic study of acute human asthma patients prior to, and following, corticosteroid treatment.* This study is not fully completed, but demonstrates the application of metabolomics towards the understanding acute human asthma pathophysiology.

Specific Aim #7 - Pediatric asthma longitudinal study

Chapter VIII – In addition to acute patients, a useful medical application of metabolomics would be the investigation of therapeutics and stabilization of human asthma over long periods of time. *This chapter documents the investigation of metabolic markers in relatively stable human pediatric asthma patients over repeat visits to an outpatient clinic.* This study attempts to identify and follow metabolic indices of stable human asthma following different therapeutic interventions and over extended periods of time.

Specific Aim #8 – Recommendations for future metabolomic studies

Chapter IV – Metabolomics has received growing interest over the last few years. Its application has included studies of simple metabolic pathways to complex disease diagnosis. *This chapter reviews the contributing works of Chapters II-VIII to the field of metabolomics, and reflects upon many philosophical challenges and questions that arose over the duration of the research for this thesis.*

LITERATURE CITED

1. Lander, E.S., et al., Initial sequencing and analysis of the human genome. Nature, 2001. 409(6822); 860-921.
2. Affolter, M., et al., -Omics for prevention: gene, protein and metabolite profiling to better understand individual disposition to disease. Nestle Nutr Workshop Ser Pediatr Program, 2006(57); 247-50; discussion 250-5.
3. Trujillo, E., C. Davis, and J. Milner, Nutrigenomics, proteomics, metabolomics, and the practice of dietetics. J Am Diet Assoc, 2006. 106(3); 403-13.
4. Joyce, A.R. and B.O. Palsson, The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol, 2006. 7(3); 198-210.
5. Butcher, E.C., E.L. Berg, and E.J. Kunkel, Systems biology in drug discovery. Nat Biotechnol, 2004. 22(10); 1253-9.
6. de Hoog, C.L. and M. Mann, Proteomics. Annu Rev Genomics Hum Genet, 2004. 5; 267-93.
7. Lamitina, T., Functional genomic approaches in C. elegans. Methods Mol Biol, 2006. 351; 127-38.
8. Heckel, D.G., Genomics in pure and applied entomology. Annu Rev Entomol, 2003. 48; 235-60.
9. Prentice, H. and K.A. Webster, Genomic and proteomic profiles of heart disease. Trends Cardiovasc Med, 2004. 14(7); 282-8.
10. Emilien, G., et al., Impact of genomics on drug discovery and clinical medicine. QJMed, 2000. 93(7); 391-423.
11. Lockhart, D.J. and E.A. Winzeler, Genomics, gene expression and DNA arrays. Nature, 2000. 405(6788); 827-36.
12. Arrell, D.K., I. Neverova, and J.E. Van Eyk, Cardiovascular proteomics: evolution and potential. Circ Res, 2001. 88(8); 763-73.
13. Pal, C., B. Papp, and M.J. Lercher, An integrated view of protein evolution. Nat Rev Genet, 2006. 7(5); 337-48.

14. Becker, K.F., et al., Clinical proteomics: new trends for protein microarrays. Curr Med Chem, 2006. 13(15); 1831-7.
15. Nicholson, J.K. and I.D. Wilson, Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. Nat Rev Drug Discov, 2003. 2(8); 668-76.
16. Nicholson, J.K., et al., Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov, 2002. 1(2); 153-61.
17. Holmes, E., T.M. Tsang, and S.J. Tabrizi, The application of NMR-based metabonomics in neurological disorders. NeuroRx, 2006. 3(3); 358-72.
18. Mitchell, S.C. and P.L. Carmichael, Metabonomics and the endocrine system. Mol Cell Endocrinol, 2005. 244(1-2); 10-4.
19. Waters, N.J., et al., Integrated metabonomic analysis of bromobenzene-induced hepatotoxicity: novel induction of 5-oxoprolinosis. J Proteome Res, 2006. 5(6); 1448-59.
20. Coen, M., et al., An integrated metabonomic investigation of acetaminophen toxicity in the mouse using NMR spectroscopy. Chem Res Toxicol, 2003. 16(3); 295-303.
21. German, J.B., B.D. Hammock, and S.M. Watkins, Metabolomics: building on a century of biochemistry to guide human health. Metabolomics, 2005. 1(1); 3-9.
22. Griffin, J.L., Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. Curr Opin Chem Biol, 2003. 7(5); 648-54.
23. Clayton, T.A., et al., Pharmaco-metabonomic phenotyping and personalized drug treatment. Nature, 2006. 440(7087); 1073-7.
24. Walsh, M.C., et al., Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. Am J Clin Nutr, 2006. 84(3); 531-9.
25. Watson, A.D., Thematic review series: Systems Biology Approaches to Metabolic and Cardiovascular Disorders. Lipidomics: a global approach to lipid analysis in biological systems. J Lipid Res, 2006. 47(10); 2101-11.

26. Ellis, D.I. and R. Goodacre, Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst, 2006. 131(8); 875-85.

27. Wendisch, V.F., et al., Emerging Corynebacterium glutamicum systems biology. J Biotechnol, 2006. 124(1); 74-92.

28. Wendisch, V.F., M. Bott, and B.J. Eikmanns, Metabolic engineering of Escherichia coli and Corynebacterium glutamicum for biotechnological production of organic acids and amino acids. Curr Opin Microbiol, 2006. 9(3); 268-74.

29. Vimr, E.R., et al., Diversity of microbial sialic acid metabolism. Microbiol Mol Biol Rev, 2004. 68(1); 132-53.

30. Choi, Y.H., et al., NMR metabolomics to revisit the tobacco mosaic virus infection in Nicotiana tabacum leaves. J Nat Prod, 2006. 69(5); 742-8.

31. Atherton, H.J., et al., A combined 1H-NMR spectroscopy- and mass spectrometry-based metabolomic study of the PPAR-alpha null mutant mouse defines profound systemic changes in metabolism linked to the metabolic syndrome. Physiol Genomics, 2006. 27(2); 178-86.

32. Drackley, J.K., S.S. Donkin, and C.K. Reynolds, Major advances in fundamental dairy cattle nutrition. J Dairy Sci, 2006. 89(4); 1324-36.

33. Viant, M.R., Revealing the Metabolome of Animal Tissues Using 1H Nuclear Magnetic Resonance Spectroscopy. Methods Mol Biol, 2006. 358; 229-46.

34. Porterfield, D.M., Measuring metabolism and biophysical flux in the tissue, cellular and sub-cellular domains: Recent developments in self-referencing amperometry for physiological sensing. Biosens Bioelectron, 2006.

35. Forst, C.V., Host-pathogen systems biology. Drug Discov Today, 2006. 11(5-6); 220-7.

36. Sabatine, M.S., et al., Metabolomic identification of novel biomarkers of myocardial ischemia. Circulation, 2005. 112(25); 3868-75.

37. Soo, E.C., et al., Selective detection and identification of sugar nucleotides by CE-electrospray-MS and its application to bacterial metabolomics. Anal Chem, 2004. 76(3); 619-26.

38. Lee, K.M., J.H. Kim, and D. Kang, Design issues in toxicogenomics using DNA microarray experiment. Toxicol Appl Pharmacol, 2005. 207(2S); 200-208.

39. Mayeno, A.N., R.S. Yang, and B. Reisfeld, Biochemical reaction network modeling: predicting metabolism of organic chemical mixtures. Environ Sci Technol, 2005. 39(14); 5363-71.

40. Bren, L., Metabolomics: working toward personalized medicine. FDA Consum, 2005. 39(6); 28-33.

41. Griffin, J.L. and M.E. Bollard, Metabonomics: its potential as a tool in toxicology for safety assessment and data integration. Curr Drug Metab, 2004. 5(5); 389-98.

42. Mathers, J.C., Plant foods for human health: research challenges. Proc Nutr Soc, 2006. 65(2); 198-203.

43. Brindle, J.T., et al., Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med, 2002. 8(12); 1439-44.

44. Chernov, M.F., et al., Proton magnetic resonance spectroscopy (MRS) of metastatic brain tumors: variations of metabolic profile. Int J Clin Oncol, 2006. 11(5); 375-384.

45. Tassenoy, A., et al., Demonstration of tissue alterations by ultrasonography, magnetic resonance imaging and spectroscopy, and histology in breast cancer patients without lymphedema after axillary node dissection. Lymphology, 2006. 39(3); 118-26.

46. Li, X.M., et al., The chinese herbal medicine formula MSSM-002 suppresses allergic airway hyperreactivity and modulates TH1/TH2 responses in a murine model of allergic asthma. J Allergy Clin Immunol, 2000. 106(4); 660-8.

47. Beasley, R., et al., Prevalence and etiology of asthma. J Allergy Clin Immunol, 2000. 105(2 Pt 2); S466-72.

48. Canadian Lung Association, 2004.

49. Glinski, M. and W. Weckwerth, The role of mass spectrometry in plant systems biology. Mass Spectrom Rev, 2006. 25(2); 173-214.

50. Smedsgaard, J. and J. Nielsen, Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. J Exp Bot, 2005. 56(410); 273-86.

51. Villas-Boas, S.G., et al., Mass spectrometry in metabolome analysis. Mass Spectrom Rev, 2005. 24(5); 613-46.

52. Guttman, A., M. Varoglu, and J. Khandurina, Multidimensional separations in the pharmaceutical arena. Drug Discov Today, 2004. 9(3); 136-44.

53. Szpunar, J., Advances in analytical methodology for bioinorganic speciation analysis: metallomics, metalloproteomics and heteroatom-tagged proteomics and metabolomics. Analyst, 2005. 130(4); 442-65.

54. Kanani, H.H. and M.I. Klapa, Data correction strategy for metabolomics analysis using gas chromatography-mass spectrometry. Metab Eng, 2006.

55. Koek, M.M., et al., Microbial metabolomics with gas chromatography/mass spectrometry. Anal Chem, 2006. 78(11); 3839.

56. Koek, M.M., et al., Microbial metabolomics with gas chromatography/mass spectrometry. Anal Chem, 2006. 78(4); 1272-81.

57. Larsen, T.O., et al., Phenotypic taxonomy and metabolite profiling in microbial drug discovery. Nat Prod Rep, 2005. 22(6); 672-95.

58. Pears, M.R., et al., High resolution 1H NMR-based metabolomics indicates a neurotransmitter cycling deficit in cerebral tissue from a mouse model of Batten disease. J Biol Chem, 2005. 280(52); 42508-14.

59. Serkova, N., et al., H-NMR-based metabolic signatures of mild and severe ischemia/reperfusion injury in rat kidney transplants. Kidney Int, 2005. 67(3); 1142-51.

60. Aranibar, N., et al., Metabolomic analysis using optimized NMR and statistical methods. Anal Biochem, 2006. 355(1); 62-70.

61. Saude, E., C. Slupsky, and B. Sykes, Optimization of NMR analysis of biological fluids for quantitative analysis. Metabolomics, 2006. 2; 113-123.

62. Saude, E.J., et al., NMR analysis of neutrophil activation in sputum samples from patients with cystic fibrosis. Magn Reson Med, 2004. 52(4); 807-14.

63. Saude, E.J. and B.D. Sykes, Urine stability for metabolomic studies: effects of preparation and storage. Metabolomics, 2006. In Press.

64. Coen, M., et al., Proton nuclear magnetic resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis. Clin Infect Dis, 2005. 41(11); 1582-90.

65. Chen, M.H., et al., MRI features of an infected cephalhaematoma in a neonate. J Clin Neurosci, 2006. 13(8); 849-52.

66. Barrett, T., et al., Macromolecular MRI contrast agents for imaging tumor angiogenesis. Eur J Radiol, 2006.

67. Thompson, R.B. and E.R. McVeigh, Cardiorespiratory-resolved magnetic resonance imaging: Measuring respiratory modulation of cardiac function. Magn Reson Med, 2006.

68. Carvalho Neto, A.D., E.L. Gasparetto, and I. Bruck, Subependymal giant cell astrocytoma with high choline/creatine ratio on proton MR spectroscopy. Arq Neuropsiquiatr, 2006. 64(3b); 877-880.

69. Woods, J.C., et al., Hyperpolarized (3)He diffusion MRI and histology in pulmonary emphysema. Magn Reson Med, 2006.

70. Plymoth, A., et al., Human bronchoalveolar lavage: biofluid analysis with special emphasis on sample preparation. Proteomics, 2003. 3(6); 962-72.

71. Loehe, F., et al., Tissue damage of non-heart-beating donor lungs after long-term preservation: evaluation of histologic alteration, bronchoalveolar lavage, and energy metabolism. Shock, 2002. 17(6); 502-7.

72. Connett, G.J., Bronchoalveolar lavage. Paediatr Respir Rev, 2000. 1(1); 52-6.

73. Rogers, G.B., et al., Bacterial activity in cystic fibrosis lung infections. Respir Res, 2005. 6; 49.

74. Bernhard, W., et al., Mass spectrometric analysis of surfactant metabolism in human volunteers using deuteriated choline. Am J Respir Crit Care Med, 2004. 170(1); 54-8.

75. Holmes, E., et al., Metabolic profiling of CSF: Evidence that early intervention may impact on disease progression and outcome in schizophrenia. PLoS Med, 2006. 3(8).

76. Zaleska, M.M. and P.K. Gessner, Metabolism of [14C]paraldehyde in mice in vivo, generation and trapping of acetaldehyde. J Pharmacol Exp Ther, 1983. 224(3); 614-9.

77. Griese, M., P. Latzin, and J. Beck, A noninvasive method to collect nasally exhaled air condensate in humans of all ages. Eur J Clin Invest, 2001. 31(10); 915-20.

78. Tsang, T.M., et al., Metabolic profiling of plasma from discordant schizophrenia twins: correlation between lipid signals and global functioning in female schizophrenia patients. J Proteome Res, 2006. 5(4); 756-60.

79. Targher, G., et al., Serum 25-hydroxyvitamin D3 concentrations and carotid artery intima-media thickness among type 2 diabetic patients. Clin Endocrinol (Oxf), 2006. 65(5); 593-597.

80. Wiberg, B., et al., Metabolic risk factors for stroke and transient ischemic attacks in middle-aged men. A Community-Based Study With Long-Term Follow-Up. Stroke, 2006.

81. Stella, C., et al., Susceptibility of Human Metabolic Phenotypes to Dietary Modulation. J Proteome Res, 2006. 5(10); 2780-2788.

82. Prietsch, V., et al., Emergency management of inherited metabolic diseases. J Inherit Metab Dis, 2002. 25(7); 531-46.

83. Barregard, L., et al., Experimental exposure to wood-smoke particles in healthy humans: effects on markers of inflammation, coagulation, and lipid peroxidation. Inhal Toxicol, 2006. 18(11); 845-53.

84. Newton, D., Human biokinetics of inhaled terbium oxide. Radiat Prot Dosimetry, 2003. 106(1); 53-61.

85. Rabinovitch, N., L. Zhang, and E.W. Gelfand, Urine leukotriene E4 levels are associated with decreased pulmonary function in children with persistent airway obstruction. J Allergy Clin Immunol, 2006. 118(3); 635-40.

86. Priftis, K.N., et al., Increased glycosaminoglycans in the urine of asthmatic children on inhaled corticosteroids. Pediatr Allergy Immunol, 2006. 17(3); 194-8.

87. Colgan, R., et al., Asymptomatic bacteriuria in adults. Am Fam Physician, 2006. 74(6); 985-90.

88. Gupta, A., et al., (1)H NMR spectroscopy in the diagnosis of Klebsiella pneumoniae-induced urinary tract infection. NMR Biomed, 2006.

89. Hazan, Z., et al., Effective prevention of microbial biofilm formation on medical devices by low energy surface acoustic waves. Antimicrob Agents Chemother, 2006.

90. de Allori, M.C., et al., Antimicrobial resistance and production of biofilms in clinical isolates of coagulase-negative Staphylococcus strains. Biol Pharm Bull, 2006. 29(8); 1592-6.

91. Kunin, C.M., Urinary-catheter-associated infections in the elderly. Int J Antimicrob Agents, 2006. 28 Suppl 1; S78-81.

92. Lin, Y.T., et al., Asymptomatic bacteriuria among the institutionalized elderly. J Chin Med Assoc, 2006. 69(5); 213-7.

93. Miles, K.I. and M.W. Wren, Evaluation of three commercial agar preparations for the presumptive identification of significant urinary isolates. Br J Biomed Sci, 2005. 62(4); 179-81.

94. Shopova, E., A. Nikolov, and A. Dimitriv, [Link between the state of vaginal flora and the development of uroinfection during pregnancy]. Akush Ginekol (Sofiia), 2005. 44(1); 38-9.

95. Geigy, a.L., Geigy scientific tables. 8th, rev. C. Lentner. Vol. 1. 1981 - 1992, West Cadwell, NJ: Ciba-Geigy Corp.

96. Gould, S.J., I havelanded: the end of a beginning in natural history. 2003, Newy York: Three Rivers Press.

97. Daffertshofer, A., et al., PCA in studying coordination and variability: a tutorial. Clinical Biomechanics, 2004. 19(4); 415-428.

98. Stoyanova, R. and T.R. Brown, NMR spectral quantitation by principal component analysis. NMR Biomed, 2001. 14(4); 271-7.

99. Bezabeh, T., et al., The use of 1H magnetic resonance spectroscopy in inflammatory bowel diseases: distinguishing ulcerative colitis from Crohn's disease. Am J Gastroenterol, 2001. 96(2); 442-8.

100. Mountford, C.E., et al., Diagnosis and prognosis of breast cancer by magnetic resonance spectroscopy of fine-needle aspirates analysed using a statistical classification strategy. Br J Surg, 2001. 88(9); 1234-40.

101. Staib, A., et al., Disease pattern recognition testing for rheumatoid arthritis using infrared spectra of human serum. Clin Chim Acta, 2001. 308(1-2); 79-89.

CHAPTER II

Optimization of NMR qualitative and quantitative analysis of body fluids

OVERVIEW

With the growing interest in the use of NMR spectroscopy for the study of biological fluids such as urine and serum for metabolomic or diagnostic purposes, new challenges have arisen concerning the efficacy of NMR data acquisition and analysis[1, 2]. In particular, the quantification of sample constituents such as metabolites is of great importance. *This chapter reviews a study that compared five one-dimensional proton NMR pulse sequences using synthetic urine samples to determine appropriate acquisition parameters for reasonable sample throughput and accuracy.* Each pulse sequence that was chosen had its own advantages and limitations with respect to solvent suppression, baseline stability, exchangeable protons, and quantization of resonances near the residual water peak.

Issues pertaining to spectrometer hardware (low-pass filters), the pulse sequence used to collect the FID, the software to analyze the spectra, as well as NMR phenomena such as longitudinal relaxation impacted the accuracy and efficacy of the qualitative and quantitative analysis of urine samples. The optimization of analysis is extremely important to ensure that the data collected and the scientific conclusions are accurate.

INTRODUCTION

Today, metabolomic and magnetic resonance diagnostic (MRD) investigations are beginning to harness the versatility of NMR to identify compounds from biological fluids such as urine, serum, or cerebral spinal fluid[1-5]. With a shift in focus from structural elucidation to more physiological questions arising from *in vivo* samples there is a need to review the efficacy of NMR data acquisition and the ability to identify and quantitate sample constituents that are thought to play a role in various pathophysiological states[6-8].

Concomitant with the use of NMR for metabolic investigations, sample numbers have also increased. Animal or clinically based studies require the analysis of multiple samples to produce statistically significant results. As such, investigators require pulse sequences that provide flat baselines, optimum solvent suppression, superior signal-to-noise, and resolution while remaining robust enough to analyze hundreds of samples without requiring lengthy pre-acquisition or pulse sequence set-up[9]. Our investigation of the efficacy of NMR qualitative and quantitative analysis focuses on synthetic urine samples analyzed by five widely used pulse sequences for the acquisition of NMR data[3, 10, 11]. The data was then analyzed by spectral integration using standard NMR software, as well as the Chenomx NMR Suite software, which is designed for analysis of NMR metabolomics data.

This chapter outlines a study I conducted that demonstrates the impact different one-dimensional ^{1}H NMR pulse sequences and acquisition parameters have upon the efficacy of solvent suppression, baseline stability, exchangeable protons, and the accuracy of quantitative analysis. The effects of different acquisition parameters are of particular importance when determining concentrations relative to the integration of internal standards and the knowledge of relaxation rates (T_1) of the nuclei being studied. This study also addressed the impact spectrometer hardware, such as low-pass filters, had upon quantitation.

EXPERIMENTAL PROCEDURES

II-A. Sample preparation

A synthetic urine buffer was used which provides a solution with similar chemical characteristics to that of real urine, allows for a known and reproducible buffering system, and has a defined metabolic profile (408 mM urea, 77.9 mM NaCl, 20.5 mM KH_2PO_4, 16.0 mM Na_2SO_4, 18.6 mM NH_4Cl, 21.2 mM KCl, 0.02%w/v NaN_3, 3.26 mM $CaCl_2$, 3.17 mM $MgCl_2$*$6H_2O$, 5% D_2O, and 0.49 mM DSS (disodium-2, 2-dimethyl (2-silapentane-5-sulphonate))[12]. Two samples containing specific chemical compounds were prepared in the laboratory using the synthetic urine buffer. Compounds included in the two samples were chosen to provide a range of NMR spectroscopic characteristics such as differing T_1's, signal multiplicities, and resonant frequencies. Standard Sample 1 contained final concentrations of 2-oxoisocaproic acid (5.02 mM), formic acid (3.78 mM), 2-oxoglutaric acid (4.95 mM), sodium acetate (1.06 mM), fumaric acid (2.99 mM), threonine (4.02 mM), and phenylalanine (4.01 mM). Standard Sample 2 contained final concentrations of fucose (10.1 mM), citric acid (3.08 mM), glutaric acid (2.15 mM), hippuric acid (7.07 mM), hypoxanthine (2.15 mM), and dimethylamine (3.19 mM). Both samples were brought to volume in the synthetic urine buffer solution (see above). All chemicals were purchased from Sigma-Aldrich (Mississauga, ON). The final pH for Samples 1 and 2 were 4.66 and 5.73, respectively. Aliquots of the two samples (600 μl) were then transferred to standard 5mm glass NMR tubes (Wilmad, NJ, USA).

A third sample was made from an amino acid standard ampoule (17 amino acids) purchased from the National Institute of Standards and Technology (SRM 2389, Standard Reference Materials, Gaithersburg, MD, USA). The 1ml ampoule was combined with 2.5 ml of synthetic buffer solution (see above) and had a final pH of 5.94. A 600 μl aliquot was transferred to a standard 5 mm glass NMR tube.

II-B. NMR Spectral Acquisition

NMR spectra were acquired on a 600 MHz Varian Inova spectrometer equipped with a 5 mm triple resonance probe with z-axis pulsed field gradients. One-dimensional ^{1}H spectra were collected at 25°C with a spectral width of 7200

Hz, 4 steady state scans preceding acquisition, and 32 transients were acquired for each spectrum which were then apodized with an exponential decay corresponding to a line broadening of 0.5 Hz prior to Fourier Transformation. Spectra were zero-filled to 128k. The acquisition time per scan was 4s, with a varying preacquisition delay of about 1s to ensure the total time per scan was 5s for all of the pulse sequences. The total time per scan of 5s was chosen to balance sample throughput for metabolomic studies as well as relative accuracy of quantitation of metabolites. The five different pulse sequences were chosen from the Varian BioPack software package (circa November 2003), but are commonly used in all NMR laboratories and are available in all manufacturers' software. All spectra were acquired in succession to minimize any spectrometer variations. For each spectrum the shimming was optimized so that the DSS linewidth was below 0.9 Hz. The first pulse sequence used a WET type of water suppression (shaped pulses and gradients) followed by a final 90° read pulse[13, 14]. The tnnoesy pulse sequence is a one-dimensional NOESY with transmitter saturation during the preacquisition delay and mixing time[15]. The tpresat sequence uses a transmitter presaturation, followed by a 90° read pulse[16]. The s2pul pulse sequence uses the decoupler for presaturation, followed by a 90° read pulse[9]. Finally, the cpmgt2 sequence uses decoupler presaturation followed by a Carr-Purcell-Meiboom-Gill spin-echo pulse train[10, 11]. The optimized 90° pulse width was 7.3μs in all of the NMR experiments. The NMR pulse sequences that used presaturation had an effective γB_1 field of 60 Hz for 900ms. WET had a z-gradient level of 30G/cm for 2ms. CPMGT2 had a 5G/cm gradient for 1ms, a spacing of 540ms between 180° pulses, and an overall tau of 80ms. Prior to acquisition each pulse sequence was optimized to ensure the best spectra were collected. Figure II-1 shows a graphical representation of the pulse sequences.

Samples 1 and 2 were re-acquired once a week for 3 weeks. To limit sample degradation, sodium azide was added to each sample (see sample preparation section above) and samples were stored in a refrigerator (4°C) between acquisitions.

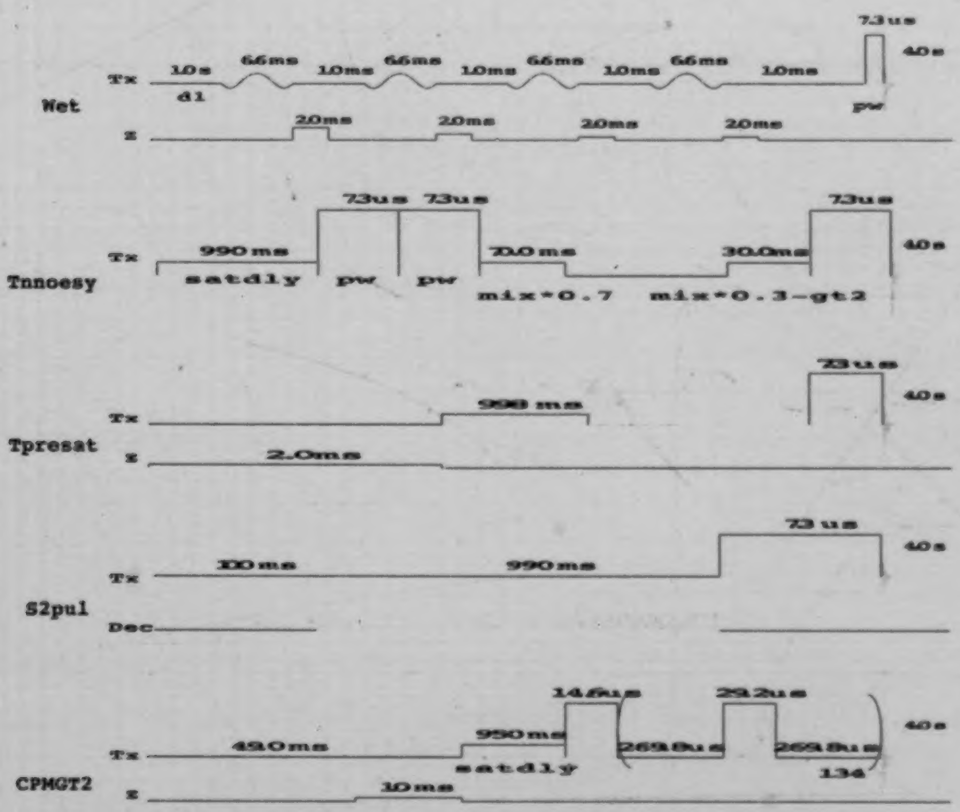

Wet
Tx
1.0s
d1
66ms
1.0ms
66ms
1.0ms
66ms
1.0ms
66ms
1.0ms
73us
pw
4.0s
z
2.0ms
2.0ms
2.0ms
2.0ms
Tnnoesy
Tx
990 ms
satdly
73us
pw
73us
pw
70.0 ms
mix*0.7
30.0ms
mix*0.3-gt2
73us
4.0s
Tpresat
Tx
998 ms
73us
4.0s
z
2.0ms
S2pul
Tx
100 ms
990ms
73 us
4.0s
Dec
CPMGT2
Tx
490ms
950ms
satdly
146us
268.8us
292us
268.8us
134
4.0s
z
10ms

Figure II-1. **Schematic of five different 1D ^{1}H-pulse sequences.**

II-C. NMR Quantification

The buffer constituent DSS served as internal standard for spectral quantification. The methyl resonance peak of DSS was utilized as an internal reference set to 0ppm. Spectral quantification used both the VNMR 6.1C software package (Varian Inc, Palo Alto, CA) and the Chenomx NMR Suite Professional software package Version 3.1 (database available at pH 7.0, Chenomx Inc., Edmonton, AB). Spectral baselines were corrected using the 'dc' macro in VNMR 6.1C, which sets the extreme 10% regions on either side of the full spectrum to zero. Areas for spectral integration were chosen, as much as possible, to include resonant peaks of interest while excluding surrounding signals[8]. Integration regions were maintained for data collected from differing pulse sequences. NMR determined concentrations were confirmed by amino acid analysis of Samples 1 and 3.

II-D. T_1 measurement

Longitudinal relaxation rates for the chemical constituents of all the samples were determined by an inversion – recovery experiment using the s2pul pulse sequence. The relaxation delay in the middle of the pulse sequence was arrayed from 0.1 to 20s, with a preacquisition delay of 25s before each transient to ensure all magnetization returned to equilibrium prior to subsequent scans.

II-E. Filter attenuation mapping

A series of spectra were collected on Sample 1 where the only differing variable during spectral acquisition was the sweep width (sw) in order to test the different effects of attenuation observed by different compounds as a result of their chemical shift and relative distance from the edge of the spectrum. One-dimensional ^{1}H-NMR spectra were acquired on a 600 MHz Varian Unity spectrometer using the s2pul pulse sequence, a temperature of 25°C, 4 steady-state scans preceding acquisition, and 32 transients. Each spectrum was apodized with an exponential decay corresponding to a line broadening of 0.5 Hz prior to Fourier Transformation.

To map the regions and the degree of signal attenuation a doped 4 Hz linewidth H_2O/D_2O (1% H_20) sample was used with 0.1mg/ml $GdCl_3$ and 0.1% DSS in a sealed 5mm NMR tube (Varian Inc, Palo Alto, CA). A one-dimensional ^{1}H spectrum was collected using the s2pul pulse sequence (saturation turned off), at 25°C, a sweep width of 7200 Hz, and the residual water peak (at the centre of the spectrum) was slowly moved to the edge of the spectrum. The experiment had an initial delay of 10s and a 21-step array of the transmitter offset, -220 Hz. to 3780 Hz. One steady state scan preceded acquisition, and one transient was acquired for each acquisition in the array, which was apodized with an exponential decay corresponding to a line broadening of 0.5 Hz prior to Fourier

Transformation. In order to compare the different low-pass filter systems of differing spectrometers the spectra were also acquired on a 600 MHz Varian "NMR system" equipped with a 5mm HX resonance probe with z-axis pulsed field gradients.

RESULTS and DISCUSSION

II-A. Pulse sequence choice and optimization

In metabolomic studies high-resolution NMR must answer both qualitative and quantitative questions. It becomes important to define the abilities of NMR to accurately report metabolite concentrations. To enhance the analytical capabilities of NMR the methods of spectral acquisition and analysis must be investigated and optimized. This study compared five commonly used NMR pulse sequences to study metabolites in a synthetic urine buffer. The synthetic urine buffer provided a solution with similar chemical characteristics to that of real urine, allowed for a known and reproducible buffering system, and had a defined metabolic profile. Various NMR parameters and hardware issues were optimized to balance sample throughput for metabolomic research, as well as ensure quantitative accuracy of metabolite concentrations. The study compared quantitation using standard NMR integration and Chenomx NMR Suite software.

A choice must be made between the time required to collect the data and an acceptable signal-to-noise ratio. All spectra were acquired with 32 transients and four steady state scans. A comparison of the NMR spectra collected by the five pulse sequences is shown in Figure II-2. Differences in exchangeable protons, baseline stability, and solvent suppression were visible for spectra collected by different pulse sequences.

For the NMR analysis of biological fluids it is important to suppress as much of the water resonance as possible so that nearby metabolite peaks may be resolved[3, 9, 17, 18]. Remaining solvent signal may affect quantitation through baseline perturbation and dynamic range issues in analogue-to-digital signal conversion. Protons from the water are in high concentration (110M) when compared with the metabolites (approximately 1mM) and the remaining water signal may still have a signal that extends over a large frequency range. This results in an elevated baseline on either side of the water, which adds to peak height and resultant quantitation[19]. This is especially important in complex fluid analysis, such as urine, where multiple peaks resonate close to water and metabolic identification may only be realized given proper solvent suppression.

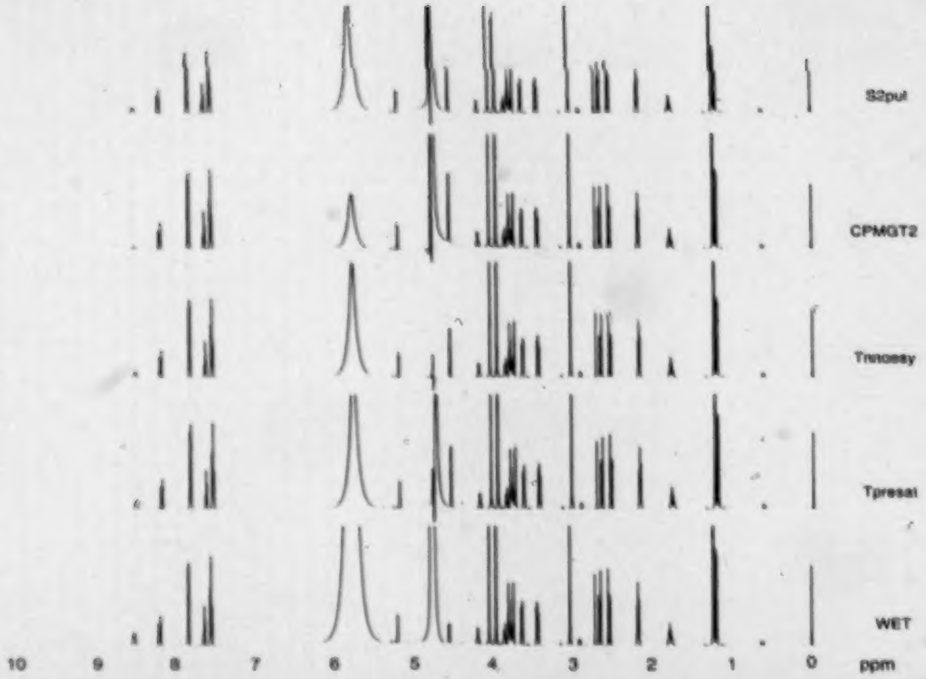


Figure II-2. **1D ^{1}H-NMR spectra acquired by different pulse sequences.** Full spectra of Sample 2 acquired by five different NMR pulse sequences. Spectral differences in baseline stability, solvent suppression, and exchangeable protons are visible.

The issue with the design and choice of appropriate NMR pulse sequences for use with water samples is maintaining acceptable solvent suppression and flat baseline in a sequence that is robust and easy to use. The five pulse sequences studied vary in efficacy of solvent suppression with the tnnoesy providing the greatest degree of water suppression, followed by tpresat as evident in the size of the remaining water signal resonating around 4.7ppm. The remaining pulse sequences were fairly similar, with tncpmgt2 the least effective (Figure II-3).

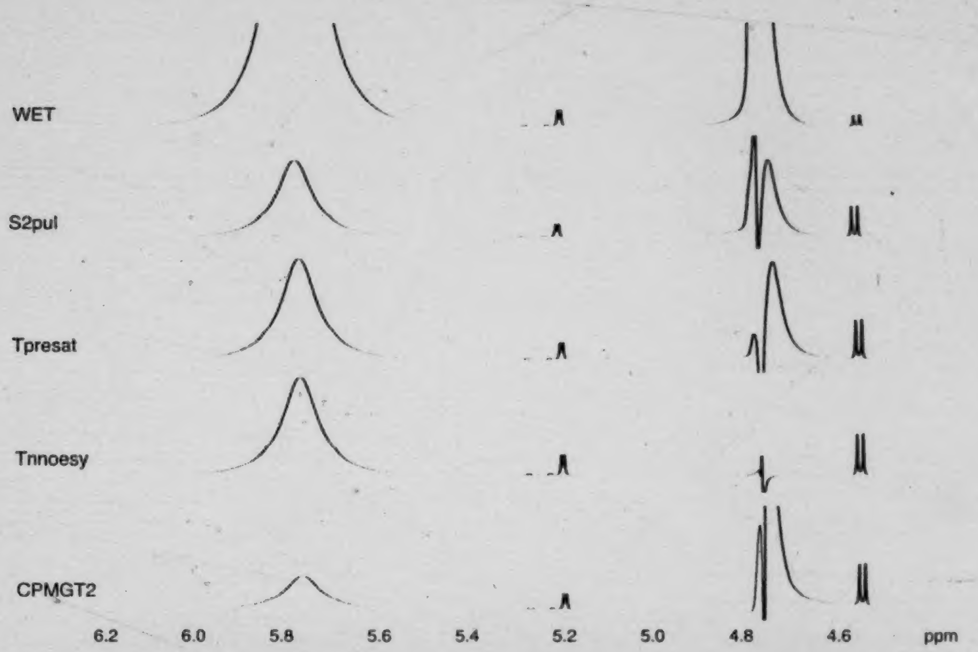


Figure II-3. **Differences in solvent suppression**. Five common 1D ^{1}H pulse sequences with different solvent suppression efficacy and baseline stability. The spectral region containing the residual water and urea peak of Sample 2, collected by the five pulse sequences. Marked differences are observed in solvent suppression, exchangeable protons, and baseline stability. The degree to which acquisition parameters and pulse sequences impact resonance quantitation is visible in the baseline elevation for the resonances from H1 (α/β) of Fucose (4.5 and 5.2ppm).

Although a few different factors may play a role in the quality of the spectral baseline the primary cause for any distortion of these samples was centered around the water and urea peaks. To quantitate baseline stability is difficult because it can be influenced by post-acquisition processing, but overall tnnoesy provided the flattest baseline across the entire spectrum. The transmitter presat pulse sequence provided a similar baseline. Spectra acquired by the s2pul and WET pulse sequences showed some distortions, especially around the water peak. Tncpmgt2 provided the lowest quality baseline due to poor solvent suppression.

The H1 proton resonance of Fucose (Sample 2), which resonates near the water, was used to test the impact of solvent suppression and the resultant baseline effects upon quantitation of nuclei. The H1 proton resonance of Fucose has two chemical shifts corresponding to the two conformations α and β. These resonances are flanked by water and urea. Tnnoesy provided the greatest degree of quantitative accuracy for these two resonant peaks with a 15% error (concentration calculated by resonance integration divided by actual concentration), followed by WET and s2pul with about 30%, tpresat had an error of 49%, and tncpmgt2 had an error of 56% (following T_1 correction, Figure II-3).

It is also important to note the influence different methods of solvent suppression play on concentration determination. The pulse sequences that saturate water (*e.g.* presat and s2pul) also attenuate protons that exchange readily with the solvent. This attenuation affects quantitation and produces inaccuracies during analysis. The samples prepared for the study were composed of a synthetic urine buffer, which contained the approximate physiological concentration of 407 mM urea. Many metabolites contain exchangeable protons. As an example, the NH_2 protons of urea are known to exchange readily with water. If the resonant peak for water is reduced during solvent suppression the resonant peak for urea can also be attenuated. Most studies do not attempt to quantitate urea so the inadvertent suppression of the urea peak may be of little consequence, other than aiding in the resolution of peaks close to the broad urea signal. However, the attenuation of urea does point to the fact that other protons will also be attenuated as a result of exchange with water. Fortunately, compounds in solution typically possess multiple resonant signals and care must be taken to report metabolic concentrations based on signals not undergoing exchange with the solvent. The pulse sequences that use water saturation inadvertently affect the urea resonance as a result of proton exchange, but because WET uses shaped pulses that leave water along the z-axis the water remains unperturbed and therefore does not affect the urea peak. Among the five pulse sequences WET produced the largest resonant peak for urea, with the remaining pulse sequences providing a relatively similar sized, and greatly attenuated, urea peak (Figure II-3).

II-B. Influence of relaxation

Acquisition time and the delays within a pulse sequence influence the total time and the recycling time allotted spin systems during acquisition. The effect of delays was tested using the s2pul experiment. The delay and acquisition time were arrayed separately (d2=1, 2, 4, and 8s; at=4, 5, and 7s), see Figure II-4. By extending the acquisition time from 4 to 7s resonance heights were significantly enhanced. The extra 3s of acquisition time allowed for an enhanced DSS methyl signal of 15% (Figure II-4). In addition, by extending the d2 delay of the s2pul pulse sequence from 1 to 8s the DSS signal improved by 18% (Figure II-5).

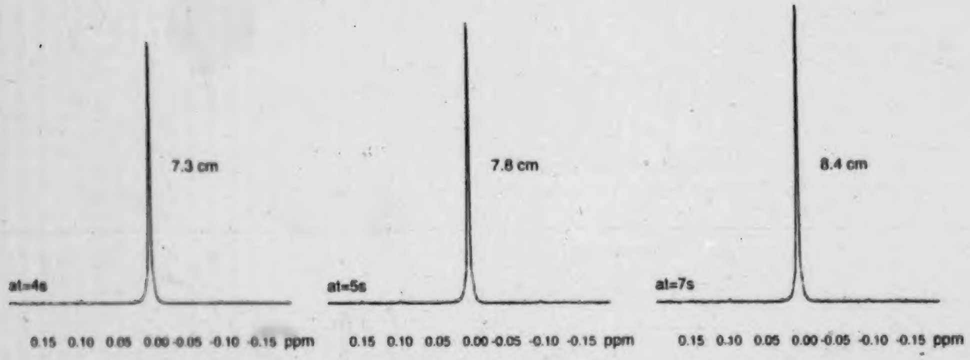


Figure II-4. **Longer acquisition time allows relaxation.** Referenced DSS resonant singlet following NMR spectral acquisition using the s2pul pulse sequence with different acquisition times. By lengthening the acquisition time, and therefore the recycling time, the DSS singlet is allowed more time to relax which is manifest by increased intensity.

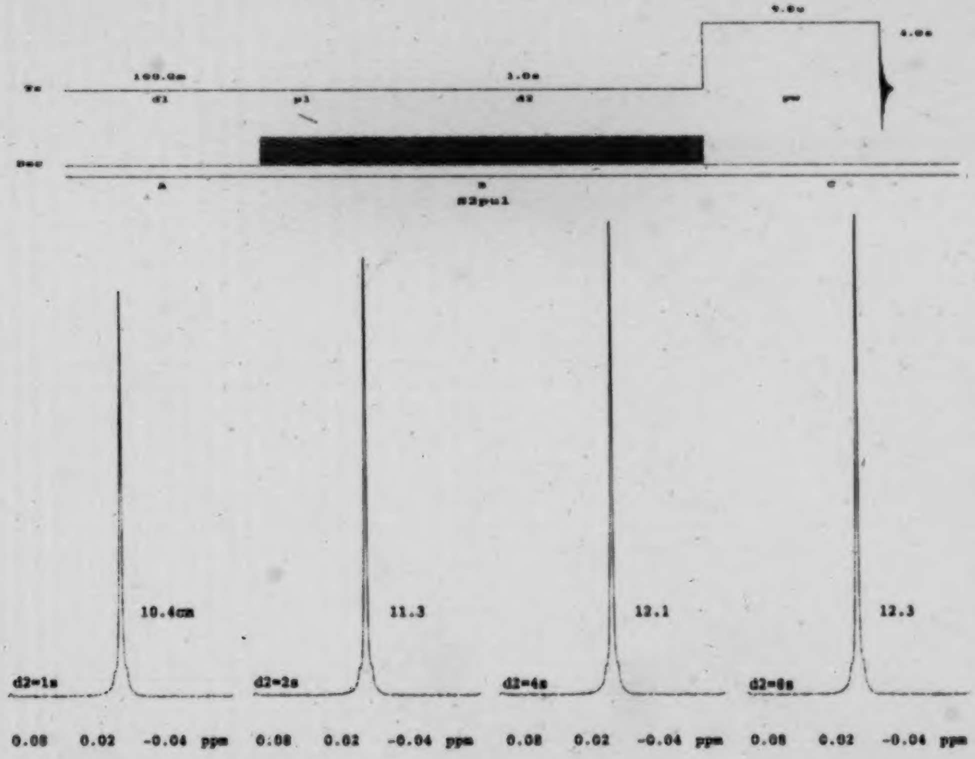


Figure II-5. **Lengthening delay time allows for fuller DSS relaxation.** Referenced DSS resonant singlet following NMR spectral acquisition using the s2pul pulse sequence with different delay times (d_2). By lengthening the delay and therefore the recycling time, the DSS singlet is allowed more time to relax before each transient, which results in an increased intensity.

Chemical concentrations were determined through integration of resonance peaks and referencing with the known internal standard DSS. It is evident that peaks from different metabolites relax and return to positive peaks at different rates. Of particular interest were the differing relaxation rates of resonant peaks originating from the same metabolite. See Figures II-6 and -7 for examples of the different relaxation rates of resonant peaks. As a result of differing T_1

delays of the chemicals found in the samples (including DSS) a T_1 correction factor was included in the study's quantitation calculations (t=d1 + at):

Eq.1
$$\frac{Int^{unknown}}{Int^{DSS}} = \frac{Conc^{unknown}[1-e^{-t/T1unknown}]}{Conc^{DSS}[1-e^{-t/T1DSS}]}$$

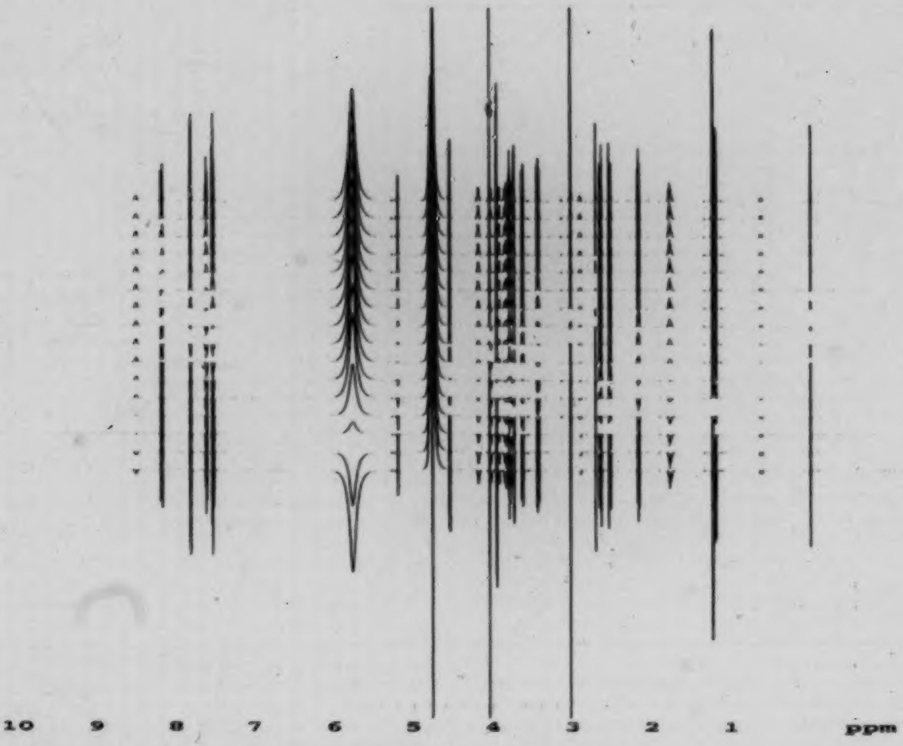


Figure II-6. **T_1 relaxation of multiple metabolites in one sample.** Stacked spectra of the resonant peaks from Sample 2 following an inversion-recovery experiment for the measurement of T_1 relaxation. The pulse sequence delay was arrayed from 0.1s to 20s; different relaxation rates are evident between the different compounds in the sample.

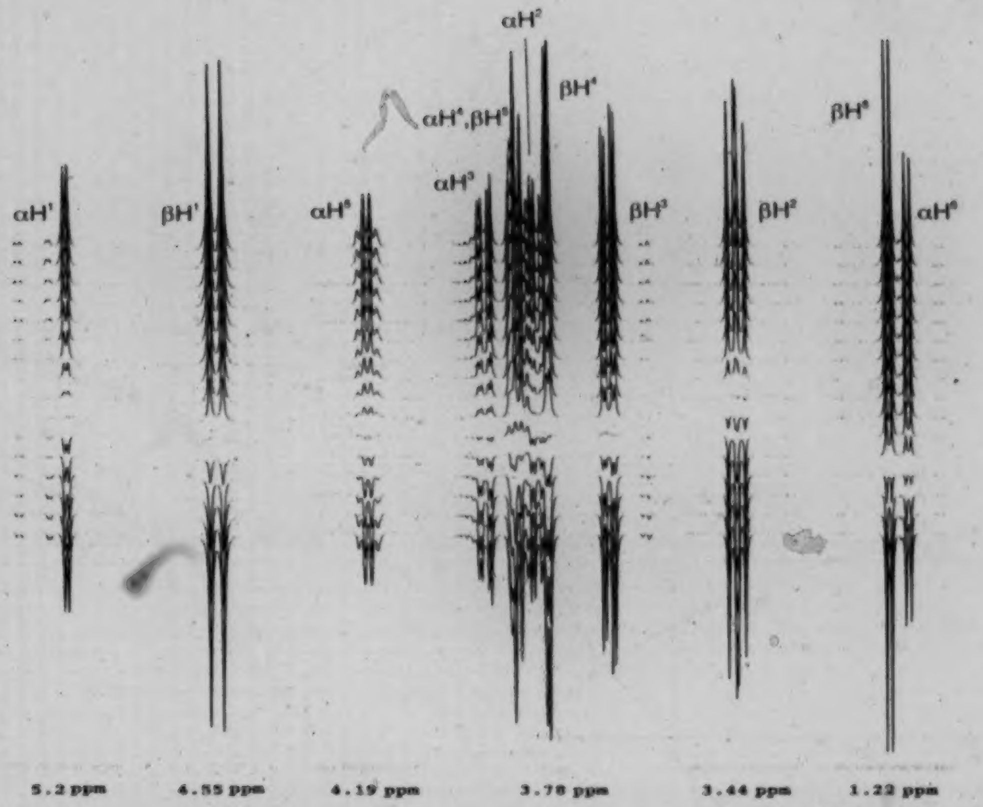


Figure II-7. **Evidence of different T_1 relaxation in one metabolite.** Stacked spectra of the resonant peaks from Fucose (Sample 2) following an inversion-recovery experiment for the measurement of T_1 relaxation. The pulse sequence delay was arrayed from 0.1s to 20s; different relaxation rates are evident from different resonant peaks of Fucose.

Evolution of magnetization following a frequency pulse may be correlated with the initial equilibrium magnetization, the exponential decay of perturbed magnetization by the time the signal is allowed to relax (5s), and the compounds longitudinal relaxation rate (T_1). The initial concentration for each compound was calculated by referencing resonant integration with that of DSS. The concentration was then multiplied by the correction factor to provide a compound specific quantitation that was not affected by differing relaxation rates. See Table II-1 for a demonstration of quantitation calculations and T_1 correction.

Compound	ppm	H's	Initial Conc(mM)[a]	T_1	T1 Corr Fact[b]	Corr Conc(mM)	Actual(mM)	Ratio[c]
Fucose	1.2/1.25	CH3 (6α)						
	1.25	CH3 (6β)	15.82	0.88	1.00	12.89	10.06	1.28
	3.4	CH (2β)						
			14.81	2.35	0.88	13.66		1.36
	3.6	CH (3β)						
			15.58	1.44	0.97	13.06		1.30
	3.74	CH (4β)						
	3.76	CH (2α)						
	3.8	CH *2 (4α,5β)						
	3.85	CH (3α)						
	4.19	CH (5a)						
	4.55	CH (1β)	15.93	1.45	0.97	13.36	10.06	1.33
	5.2	CH (1α)		2.08	0.91			
Citric Acid	2.5	CH2	4.47	0.86	1.00	3.64	3.08	1.18
	2.6	CH2		0.89	1.00		3.08	
Glutaric Acid	2.2	(CH2) * 2	3.17	1.64	0.95	2.71	2.15	1.26
Hippuric Acid	3.9	CH2	10.84	1.52	0.96	9.15	7.07	1.29
	7.6	(CH) * 2	9.71	3.25	0.79	10.04		1.42
	7.65	CH	9.07	3.66	0.75	9.88		1.40
	7.8	(CH) * 2	9.94	2.95	0.82	9.90		1.40
	8.5	NH	8.09	0.58	1.00	6.57		0.93
Hypoxanthine	8.2	(CH) * 2	2.70	4.97	0.63	3.45	2.15	1.61
Dimethylamine	2.7	(CH3) * 2	3.35	3.94	0.72	3.79	3.19	1.19

Table II-1. Metabolite concentration determination via resonance integration and T1 correction of Sample 2

a. $[Unknown] = \left(\frac{[DSS]*\#\ protons_{DSS}}{Integration_{DSS}}\right)\left(\frac{Integration_{unknown}}{\#\ protons_{unknown}}\right)$

b. See Eq. 1 in text

c. Ratio of calculated concentration/Actual concentration: expression of quantitative accuracy

Raw calculation of concentrations produced an inaccuracy ranging over differing resonant peaks from 1 – 86%, depending on the compound, resonant of choice, and pulse sequence. Over the five pulse sequences the average non-corrected accuracy error was 35%. To determine the influence of relaxation and T_1 differences inherent to the multiple small molecules found in biological samples the resonant specific T_1's were calculated for all the compounds, see Figure II-5. After incorporating the T_1 correction factor the inaccuracy over differing peaks and pulse sequences dropped to 1- 56%. Over the five pulse sequences, the average T_1-corrected concentration inaccuracy was 28%, see Table II-1.

One of the most important results of the study was the influence of relaxation upon quantitative accuracy[7]. By altering the length of the delays and acquisition time in pulse sequences the overall recycling time for the entire pulse sequence became greater. This lengthening of the recycling time allowed for greater, and in some cases, full relaxation of magnetization prior to continued pulsing. The allowance for complete relaxation following each scan allowed for the acquisition of the full resonant intensities, but required exceedingly long acquisition times for each sample and the overall study. Metabolomic studies typically require the analysis of hundreds of samples, which places limitations on collection time. As a result it is not practical to allow full relaxation of magnetization following each transient during data collection (*i.e.* five times T_1). Instead a compromise was made whereby an acceptable amount of time was allowed for the magnetization to evolve and relax during acquisition. A correction factor was included during quantitative calculations that took into account the recycling time and the differing relaxation rates between particular resonant peaks, particularly the internal standard. The study demonstrated that the T_1 correction factor improved quantitative accuracy. The inclusion of the T_1 correction factor consistently showed an improvement of at least 10% in quantitation accuracy. The T_1 values in urine may, of course, be different from those measured herein for synthetic urine, and therefore would have to be measured for the actual samples studied. For example, T1's could be shortened if paramagnetic metal ions are present. This would in fact improve accuracy for shorter spectral acquisition times.

II-C. Filter Attenuation

Spectra were collected on a 600 MHz Varian Unity spectrometer using the s2pul pulse sequence and varying only the sweep width (7200 Hz, 8000 Hz, and 9000 Hz) (Figure II-8). Ten minutes was allowed to pass between the completion of the preceding and the subsequent experiment; this was to ensure full relaxation of all compounds prior to the next acquisition. The referenced singlet peak from the methyl functional group of DSS increased as a result of changing the sweep width.

The relative attenuation also depended on the chemical shift. In Figure II-8 it is apparent that the relative increase in peak height is disproportionate. The resonance at 1.3ppm (Threonine) increased the least over the increasing sweep widths. However, the resonance at 0.9ppm (2-oxoisocaproic acid) increased by a significant amount. It becomes clear that resonant peaks at the edge of the spectrum (*e.g.* DSS) are affected to a greater extent than those more towards the middle of the spectrum (*e.g.* Threonine). Due to slight fluctuations in linewidth (inconsistencies with shimming) between the three experiments a direct comparison was not possible or accurate; however, it raised the possibility that frequency filters were disproportionately introducing resonant attenuation according to resonant frequency.

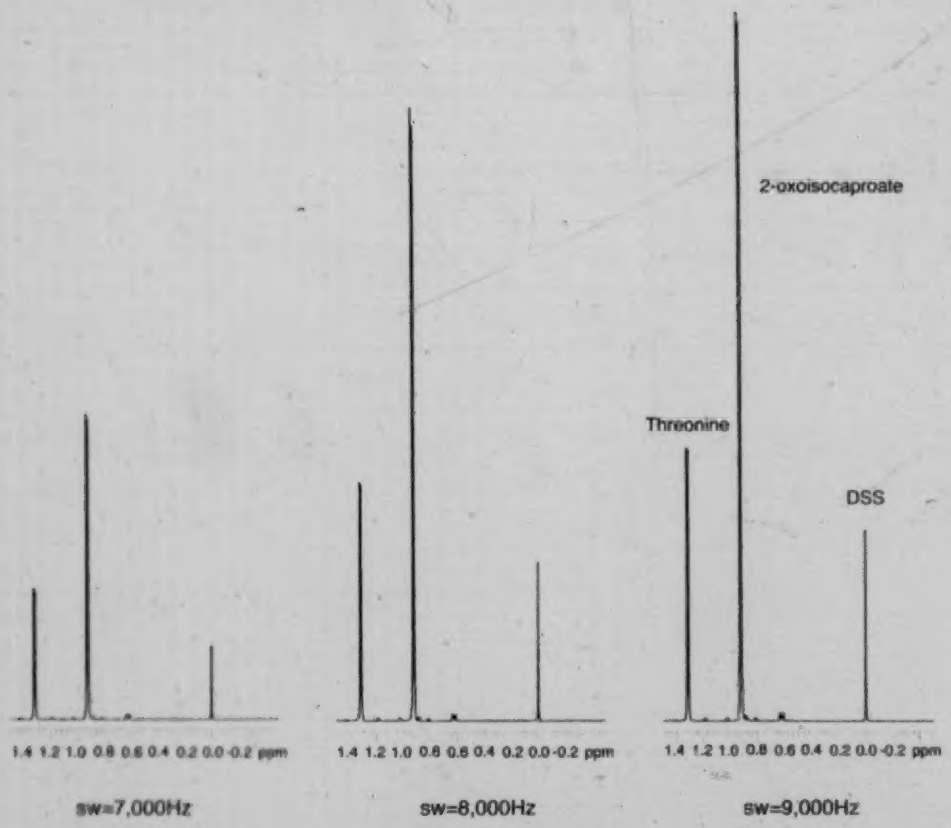


Figure II-8. **Resonance attenuation based upon chemical shift.** Aliphatic region (0-2ppm) of NMR spectra of Sample 1 collected by s2pul pulse sequence on a 600 MHz Inova spectrometer with different sweep widths. By changing only the sweep width resonant peaks increase in intensity. Resonant peaks near the center of the spectra experienced less of an increase when compared to those near the spectral edge (0ppm). Similar effects were seen in the aromatic region (8ppm) of the spectrum, indicating probable filter attenuation effects.

To map the extent and the limits of filter attenuation an experiment was conducted where a D_2O sample was analyzed and the residual water peak, which began at the centre of the spectrum, was slowly shifted to the edge of the spectrum by arraying transmitter offset. As the water peak approached the edge of the spectrum the peak height slowly deteriorated allowing for the mapping of the degree of filter attenuation relative to chemical shift. The experiment was acquired on an Inova 600 MHz spectrometer (Figure II-9) and again on a recently purchased Varian 600MHz NMR system (Figure II-10). These spectrometers have different types of filters, which had a clearly visible effect on the shape and degree of attenuation of the signals being detected and measured. The Inova 600MHz spectrometer has elliptical filters, while the Varian 600MHz NMR system has a 'brick wall' filter.

Every NMR system has unique low-pass frequency filters that remove unwanted noise from the external environment. These low-pass filters are engineered to vary along with the sweep width (region of collected frequencies) so that appropriate frequency signals are collected from the sample while disregarding additional noise. The frequency specific attenuation within the spectrum demonstrated that the further away a frequency was from the cut-off points of the sweep width and filters the less the attenuation experienced (Figure II-9). This study found that even at 3ppm there remained a 1% attenuation on peak height. By mapping the degree of attenuation at a given sweep width for different resonant frequencies I was able to implement a filter correction factor that cancelled any attenuation during post-acquisition quantitation calculation. The effect of filters may extend beyond metabolic studies when considering the effects they may have within two-dimensional experiments. In many 2D NMR experiments signal from outside the spectral window is purposefully folded into the spectrum. These frequencies may be significantly attenuated, effectively reducing the qualitative strength of two-dimensional NMR experiments.

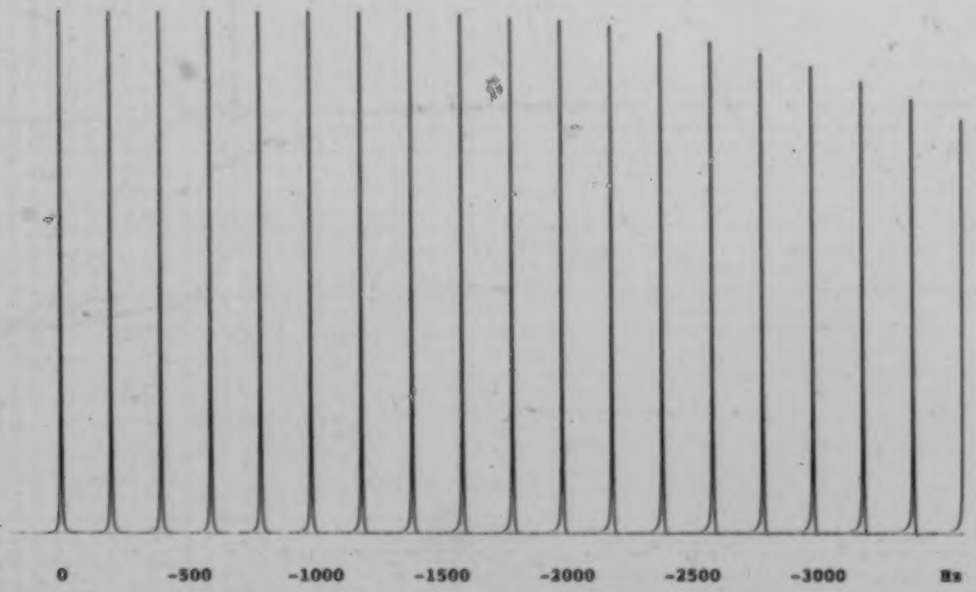


Figure II-9. **Elliptical filter attenuation of signal intensity.** Overlay of multiple NMR spectra acquired on an Inova 600 MHz magnet where the residual water peak of a deuterated sample was shifted from the center of the spectrum to the far edge by arraying the transmitter offset. As the water resonance nears the spectral edge the intensity drops indicating probable filter attenuation of resonant frequencies (elliptical phase filter). Severity of attenuation mapped to chemical shift so that a quantitation correction factor may be used to improve quantitative accuracy.

The Varian NMR system has a 'brick wall' filter, which had little to no attenuation of the DSS or other frequencies of interest (Figure II-10). Therefore spectra collected on this system are not attenuated and require no post-acquisition quantitation correction as long as the reference DSS singlet is not at the extreme spectral edge.

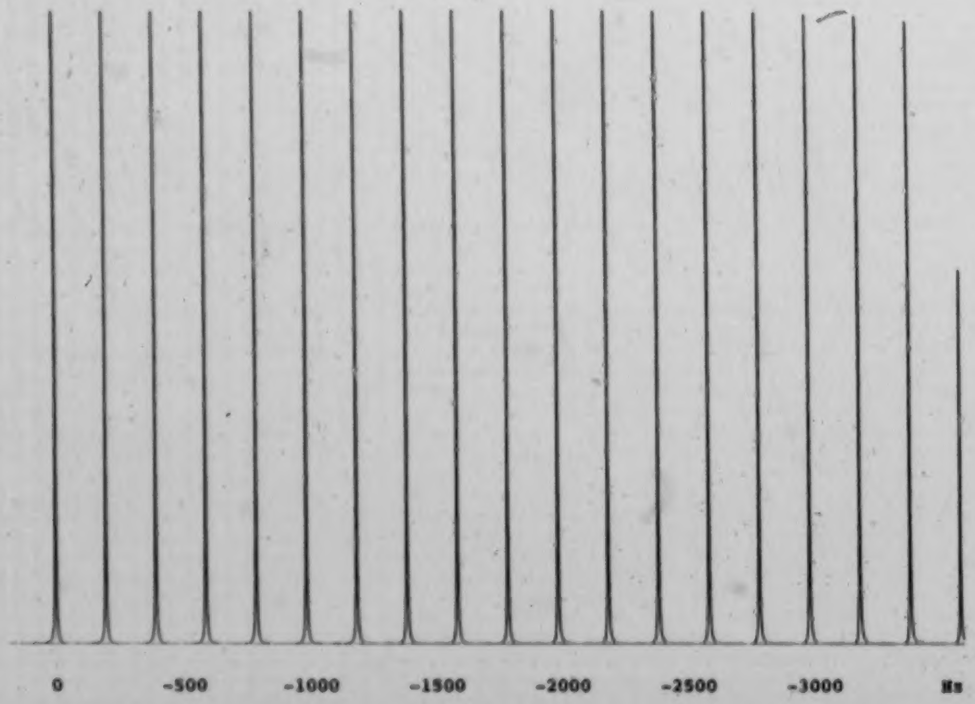


Figure II-10. **'Brick wall' filter attenuation of signal intensity.** Overlay of multiple NMR spectra acquired on a Varian 600 MHz NMR system where the residual water peak of a deuterated sample was shifted from the center of the spectrum to the far edge by arraying the transmitter offset. As the water resonance neared the spectral edge the intensity remained roughly the same, until reaching the extreme edge. The Varian 600 MHz NMR system possessed a 'brick wall' low-pass filter.

Including a correction factor for each resonance based upon its chemical shift correct for the attenuation of signals by the filters:

Eq. 2

$$CorrFact = \frac{141}{(141 - 5x10^{-7}y^2 - 2x10^{-13}y^4)}$$

Equation 2 represents a fitting of the shape of the filter attenuation of peak intensity as a function of the offset y (Hz) of the resonance from the center of the spectrum (Figure II-9). The corrected integrals are shown in Table II-2. Following filter correction quantitative accuracy averaged 15%.

Compound	ppm	H's	Filt Corr(mM)^a	T_1	T_1 Corr Fact^b	Corr Conc(mM)	Actual(mM)	Ratio^c
2-Oxoisocaproate	0.9	(CH3)2	1.06	2.07	0.91	5.63	5.13	1.10
	2.09	H	1.02	3.87	0.72	5.65		1.10
	2.6	CH2	1.01	3.79	0.73	6.42		1.25
Formic Acid	8.4	H	1.05	8.07	0.46	4.36	3.68	1.19
2-Oxoglutarate	2.4	CH2	1.01	2.05	0.91	5.34	5.05	1.06
	3	CH2	1.01	2.01	0.92	5.25		1.04
Acetic Acid	1.9	CH3	1.02	4.75	0.65	1.24	1.38	0.89
Fumaric	6.5	CH * 2	1.01	5.76	0.58	3.92	3.13	1.25
Threonine	1.3	CH3	1.04	1.10	0.99	4.61	4.07	1.13
	3.6	CH	1.00	2.43	0.87	4.56		1.12
	4.25	CH	1.00	2.23	0.89	6.53		1.60
Phenylalanine	3.1	CH	1.00	0.96	0.99	5.15	4.02	1.28
	3.3	CH	1.00	0.95	0.99	5.46		1.36
	4	CH	1.00	1.82	0.94	4.75		1.18
	7.45	CH2	1.02	3.01	0.81	4.88		1.21
	7.35	CH	1.02	3.30	0.78	5.21		1.30
	7.32	CH2	1.02	2.57	0.86	6.45		1.60

Table II-2. Metabolite concentration determination via resonance integration with T1 and filter correction for Sample 1

a. See text for filter correction, Eq.2
b. See text for T_1 correction, Eq.1
c. Ratio of calculated concentration/Actual concentration; expression of quantitative accuracy

There remains some known error in the integration as a result of peak width and the inability to fully integrate the full area of every resonance[8]. The integration package within the Varian software calculates the area under a certain peak, as specified by the user. The difficulty is that in the artificial urine samples, and certainly in rich biological spectra, resonant peaks often overlap one and another. The user may choose to include overlapping regions or to disregard them in an attempt to limit integration to spectral regions occupied by a single resonance. As an example, the internal reference singlet at 0ppm is from the methyl groups on DSS. If one looks closely at the referenced singlet further splitting and couplings are observed that arise as a result of natural abundance ^{13}C and ^{29}Si on the DSS molecule. The satellites of ^{29}Si are quite close to the DSS singlet and are often included in the integration, however the ^{13}C satellites are roughly +/-0.12 ppm away. The natural abundance of ^{13}C is roughly 1%; therefore, during analysis the integral window is chosen to be just proximal to the singlet you will knowingly lose at least 1% of the DSS integration. However, even when disregarding the ^{13}C satellites of DSS the integration window remains large (0.2ppm) when compared to regions such as around 3.7ppm where Fucose (Sample 2) has multiple resonant peaks from four different protons within a similar spectral span of 0.2ppm. During integration of resonant-rich regions the window must be kept small so that surrounding resonant peaks from different protons are not included. If the integration window is reduced to the region of DSS, where the threshold from the top of the ^{13}C satellites (0.5% each of the full DSS peak) intersects the DSS singlet, the integration is reduced to 95-96% of the full DSS peak. This is a simple example of how a 5% integration error is quickly included by simply reducing the integration window by a small amount. This loss of integration is perpetuated throughout the spectrum where integral windows must be small enough to incorporate as much of the peak as possible while limiting the inclusion of surrounding peaks. This error due to the limitations of integration window size could equate to a 10% error in concentration calculation when compounded between the resonance being quantitated and the DSS reference peak.

II-D. Amino acid analysis

Amino acid analysis was used to estimate any error that could have been attributed to inaccuracies in sample preparation. A vial containing known amino acid concentrations was purchased from the National Institute of Standards and Technology (SRM 2389, Standard Reference Materials, Gaithersburg, MD, USA), and the standard synthetic urine buffer was added. NMR quantitative analysis was limited to 15 amino acids present in solution whose resonant peaks did not overlap with surrounding signals. Among the five pulse sequences and the resonant peaks chosen the overall quantitative accuracy varied from 10-15%. Amino acid analysis of the sample (n=5) indicated a maximum 5-8% error due to sample preparation techniques.

II-E. Software

Spectra from Samples 1 and 2, acquired by the five pulse sequences, were also analyzed using the software package available from Chenomx (Chenomx Inc., Edmonton, AB, Canada). The underlying principle of the software is that experimental spectra of pure compounds are reconstructed (fit) from a linear sum of the individual spectral signatures stored in the metabolic database. Samples 1 and 2 were processed using the available software, but because the compounds were fitted as a profile including all of the resonant peaks for the metabolite, a filter correction factor could not be applied (T_1 correction is intrinsic to the specific acquisition parameters the spectral signatures were acquired under for the software database), Table II-3. Among the five different pulse sequences the quantitative inaccuracy was 10% (concentration determined by the software divided by expected concentration).

Compound	ppm	H's	Calc. Conc.	Actual Conc.	Ratio[a]
Fucose	1.2/1.25	CH3 (6a)	9.28	10.06	0.92
	1.25	CH3 (6b)			
	3.4	CH (2b)	9.28		0.92
	3.6	CH (3b)	9.28		0.92
	3.74	CH (4b)			
	3.76	CH (2a)			
	3.8	CH *2 (4a,5b)			
	3.8	H (3a)			
	4.19	H (5a)			
	4.55	H (1b)	9.28	10.06	0.92
	5.2	CH (1a)	9.28		
Citric Acid	2.5	CH2	3.33	3.08	1.08
	2.6	CH2	3.33	3.08	1.08
Glutaric Acid	2.2	(CH2) * 2	2.48	2.15	1.15
Hippuric Acid	3.9	CH2	6.71	7.07	0.95
	7.6	(CH) * 2	6.71		0.95
	7.65	CH	6.71		0.95
	7.8	(CH) * 2	6.71		0.95
	8.5	NH	6.71		0.95
Dimethylamine	2.7	(CH3) * 2	3.28	3.19	1.03

Table II-3. Concentration determination by Chenomx NMR Suite Software (Sample 2)
a. Ratio of calculated concentration/actual concentration; expression of quantitative accuracy

This study demonstrated that peak intensities differ, even within a compound, as a result of differing relaxation rates and filter attenuation. The disadvantage with the software's method of spectral fitting is that it compares the full spectral profile of the pure compound leading to some discrepancy in intensity fitting between different peaks of a given compound and therefore concentration determination. As such, a compromise had to be made among different peaks from the same compound. Often, if one resonance peak was fit another resonance for the same compound, although present at the correct frequency, was not at the correct height (due to different relaxation rates and filter attenuation of resonances within the same compound). A compromise of resonance height, and therefore concentration, had to be made among the different resonance peaks for each compound when analyzed by the Chenomx software. Since the Chenomx database is composed of previously acquired spectra of pure compounds using particular acquisition parameters, for optimum fitting and quantitative analysis subsequent spectra analyzed by the software must be collected with exactly the same parameters as those used for the database. This increases the probability of maximum compatibility.

The advantage of the Chenomx software was that if a resonance was overlapped in the spectrum being analyzed the peak height could still be estimated and a concentration determined. Alternately, in the integration software from Varian the overlapping portion of the resonance would be disregarded, the integration window reduced, and an error in quantitation would occur. Across the five different pulse sequences and the different samples analyzed, the quantitative inaccuracy for the Chenomx software remained roughly 10%.

II-F. Precision

The sections above attempt to define and optimize the quantitative accuracy of 1D-NMR (the ability to correctly describe the true concentration of a compound). However, it is also important to define the precision of NMR analysis (similar concentrations upon repeated of measurements). To separate accuracy from precision, Samples 1 and 2 were collected every week for 3 weeks on the Inova 600 MHz spectrometer. Following repeated quantitative analysis of the sequential spectra the final conclusions were identical (data not shown). The samples were also run on a Unity 600 MHz magnet and the spectra were analyzed to investigate differences between similar magnets. Following quantitative analysis the calculated metabolite concentrations were found to be identical to those from the Inova 600 MHz magnet. Due to the similar software and hardware the spectra were very similar between the two spectrometers. As well, repeated data acquisition returned nearly identical spectra, indicating spectrometer stability[19, 20].

Finally, to determine the effect of operator-set phasing upon quantitative accuracy the same spectra were analyzed, cleared, re-phased, and then analyzed again. The calculated concentrations were within 1-2% of the original analysis (data not shown).

CONCLUSION

This study demonstrates that by adopting and maintaining the NMR acquisition parameters one may now assign a well characterized confidence interval to their quantitative analysis[6, 20]. I have shown that this quantitative analysis has an accuracy limit that may be defined and will remain precise among similar magnets and spectral acquisition runs. The quantitative integration is also robust enough to handle repeat analysis and minor changes in user-set phasing. Although the tnnoesy pulse sequence provided superior water suppression and baseline, the five pulse sequences were quite similar in overall quantitative accuracy. Each of the five pulse sequences had its own advantages and disadvantages, such as solvent suppression, exchange, and baseline effects. To minimize the effect of T_1 relaxation it is best to choose a relatively long overall recycling time (*i.e.* 5s). To minimize the effect of any filter attenuation, the simplest solution is to extend the spectral width used.

My study demonstrated that the optimization of the NMR acquisition, as well as the quantitative analysis, is key for the movement of NMR to the unique field of metabolic investigations. Investigators may proceed with confidence in their analysis of biological fluids, having optimized their data acquisition and characterized the specific limitations of their quantitative accuracy.

LITERATURE CITED

1. Brindle, J.T., et al., Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med, 2002. 8(12); 1439-44.
2. Coen, M., et al., Proton nuclear magnetic resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis. Clin Infect Dis, 2005. 41(11); 1582-90.
3. Potts, B.C., et al., NMR of biofluids and pattern recognition: assessing the impact of NMR parameters on the principal component analysis of urine from rat and mouse. J Pharm Biomed Anal, 2001. 26(3); 463-76.
4. de Graaf, R.A. and K.L. Behar, Quantitative 1H NMR spectroscopy of blood plasma metabolites. Anal Chem, 2003. 75(9); 2100-4.
5. Weckwerth, W. and K. Morgenthal, Metabolomics: from pattern recognition to biological interpretation. Drug Discov Today, 2005. 10(22); 1551-8.
6. Cavaluzzi, M.J., D.J. Kerwood, and P.N. Borer, Accurate nucleic acid concentrations by nuclear magnetic resonance. Anal Biochem, 2002. 308(2); 373-80.
7. Kriat, M., et al., Quantitation of metabolites in human blood serum by proton magnetic resonance spectroscopy. A comparative study of the use of formate and TSP as concentration standards. NMR Biomed, 1992. 5(4); 179-84.
8. Griffiths, L., Assay by nuclear magnetic resonance spectroscopy:quantification limits. Analyst, 1998. 123; 1061-1068.
9. Hore, P.J., Solvent Suppression in Fourier Transform Nuclear Magnetic Resonance. Journal of Magnetic Resonance, 1983. 55; 283 - 300.
10. Van, Q.N., G.N. Chmurny, and T.D. Veenstra, The depletion of protein signals in metabonomics analysis with the WET-CPMG pulse sequence. Biochem Biophys Res Commun, 2003. 301(4); 952-9.

11. Lenz, E.M., et al., A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. J Pharm Biomed Anal, 2003. 33(5); 1103-15.

12. Jones, D.S., Djokic, J., and Gorman, S.P., Characterization and optimization of experimental variables within a reproducible bladder encrustation model and in vitro evaluation of the efficacy of urease inhibitors for the prevention of medical device-related encrustation. Journal of Biomedical Materials Research Part B: Applied Biomaterials, 2005. 76B(1); 1 - 7.

13. Smallcombe, S.H., Patt, S. L., and Keifer, P. A., WET Solvent Suppression and Its Applications to LC NMR and High-Resolution NMR Spectroscopy. Journal of Magnetic Resonance, 1995. Series A 117; 295 - 303.

14. Ogg, R.J., P.B. Kingsley, and J.S. Taylor, WET, a T1- and B1-insensitive water-suppression method for in vivo localized 1H NMR spectroscopy. J Magn Reson B, 1994. 104(1); 1-10.

15. Kumar, A., R.R. Ernst, and K. Wuthrich, A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem Biophys Res Commun, 1980. 95(1); 1-6.

16. Hoult, D.I., Solvent peak saturation with single phase and quadrature Fourier transformation. Journal of Magnetic Resonance, 1976. 21; 337 - 47.

17. Sykes, B.D., S.L. Patt, and D. Dolphin, The role of distortion in the lysozyme mechanism. Cold Spring Harb Symp Quant Biol, 1972. 36; 29-33.

18. Hwang, T.L., and A.J. Shaka, Water Suppression That Works. Excitation Sculpting using Arbitrary Waveforms and Pulsed Field Gradients. Journal of Magnetic Resonance, 1995. Series A(112); 275-79.

19. Keun, H.C., et al., Analytical reproducibility in (1)H NMR-based metabonomic urinalysis. Chem Res Toxicol, 2002. 15(11); 1380-6.

20. Crockford, D.J., et al., Curve-fitting method for direct quantitation of compounds in complex biological mixtures using 1H NMR: application in metabonomic toxicology studies. Anal Chem, 2005. 77(14); 4556-62.

CHAPTER III

Urine as a biofluid for metabolomic studies

OVERVIEW

Metabolomic studies attempt to identify and profile unique metabolic differences among test populations, which may be correlated to a specific pathophysiology. Due to the ease of collection and the metabolite-rich nature of urine, it is frequently used as a bio-fluid for human and animal metabolic studies. However, the method of urine collection is often restricted in human clinical studies due to the location of the individual, the hospital or clinic, and the immediacy of medical care. It is important to identify and define proper sample handling and storing procedures to ensure the sample has been preserved appropriately. *This chapter describes my investigation into the impact preparatory and storage procedures have on urine metabolites.*

Urine was collected from a relatively healthy male and female subject. The urine was prepared: raw, following centrifugation, filtration, or the addition of the bacteriostatic preservative sodium azide, and analyzed by NMR. Triplicates of each preparatory step were stored at room temperature (22°C), in a refrigerator (4°C), or in a deep-freeze (-80°C). Metabolites that were easily identified and commonly reported in other metabolic studies were quantified and followed over 30 days. Urine samples used in clinical and metabolomic studies are often collected and allowed to sit at room temperature for extended periods of time. In addition, the samples usually remain in a freezer until enough samples accumulate to begin to prepare and acquire the NMR spectra.

INTRODUCTION

Metabolomic investigations attempt to detect and profile fluctuations in metabolites, which may reflect changes in metabolic pathways with relation to a certain disease [1, 2]. To follow human metabolic pathways, biological tissues or fluids are sampled from the subject or patient, and analyzed for changes in metabolite concentrations from baseline levels.

Currently the metabolic analysis of a whole human subject is not possible so a sample of tissue or biofluid is used. Studies have analyzed many different biological fluids, including urine, sputum, bronchoalveolar lavage, breath condensate, whole blood, serum, feces, and cerebral spinal fluid [3, 4]. Urine is a popular biofluid for metabolomic investigations due to the opportunity for non-invasive collection, the complex metabolic nature of the fluid, and the ability to collect multiple samples over a period of time.

The detection of metabolites and the information gained by tracking metabolic flux provides information regarding an animals' physiology or disease state. Many different methods of detection have been used for metabolomic studies, such as high performance liquid chromatography (HPLC), high-resolution

NMR spectroscopy, and mass spectrometry (MS)[2, 5, 6]. Recent technological advancements and studies have enhanced and defined the ability of ^{1}H-NMR to accurately profile the metabolic makeup of a biofluid qualitatively and quantitatively[7].

The correlation of metabolite changes with a particular physiology is the basis for metabolomic investigations. Many studies have found correlations of key metabolites with particular diseases, metabolites that may be used for diagnosis, or changes that may arise due to laboratory conditions of the study [8-10]. Human urine studies have found changes in metabolite concentrations as a result of age and diet, variation in urine metabolite concentrations within large populations, and variation of particular urine metabolites in individuals over time [3, 11-16]. Studies of animal urine have found metabolic differences within genetically similar animal models[17, 18]. In addition, studies have found metabolic changes following the movement of animals from a sterile environment to one that is open to ambient pathogens[19]. To ensure the fidelity of the biological sample it is clear that attention must be given to how a sample is prepared and stored prior to analysis.

This chapter documents a study into the effect sample handling has upon the metabolic make-up of urine. Different methods of sample preparation (centrifugation, filtration, or addition of the preservative sodium azide), as well as sample storage (room temperature (22°C), refrigerator (4°C), or the deep-freeze (-

80°C)) have significant effects upon the changes of urine metabolites over time, as detected by ^{1}H-NMR. By outlining proper sample handling and storage techniques for urine samples further metabolomic studies may ensure that the sample reflects the original metabolic state of the subject.

EXPERIMENTAL PROCEDURES

III-A. Urine Collection

One healthy male and female volunteer provided urine samples in accordance with the guidelines established by the University of Alberta Health Research Ethics Board. Mid-stream urine samples were collected in the morning in a sterile container and immediately prepared as outlined below.

III-B. Sample Preparation

Sample preparation was duplicated for both male and female urine samples and in triplicate for the three different methods of storage. In a biosafety fume hood the fresh urine samples were prepared by transferring a 630 μl aliquot of urine to a 1.5 ml Eppendorff tube followed by the addition of 70 μl of a standard solution (4.9 mM DSS (disodium-2, 2-dimethyl 2-silapentane-5-sulphonate), and 100.0 mM imidazole in D_20, Sigma-Aldrich, Mississauga, ON). Filtered urine samples were prepared by filtering 1 ml of urine through a 0.22 μm syringe filter (Millipore, Cambridge, ON,). A 630 μl aliquot of the filtered urine was transferred to a new Eppendorff tube and 70 μl of the standard solution was added (see above). Spun urine samples were prepared by the centrifugation of 1ml of urine in an Eppendorff tube at 10,000 rpm for 10 min (MSE Microcentaur, Sanyo-Gallenkamp). A 630 μl aliquot of the spun urine was placed in a new Eppendorff followed by the addition of 70 μl of the standard solution (see above). For urine samples with the preservative sodium azide, 1 ml of raw urine was transferred to a 1.5 ml Eppendorff tube followed by the addition of a stock solution of sodium azide (Sigma Aldrich) to reach a final sample concentration of 0.1, 1.0, and 10 mM sodium azide. A 630 μl aliquot of the urine and sodium azide solution was transferred to a 1.5 ml Eppendorff tube followed by the addition of 70 μl of the standard solution (see above). For each of the urine samples (fresh, spun, filtered, and those containing sodium azide) a final aliquot of 600 μl was transferred to a standard 5 mm glass NMR tube (Wilmad, NJ, USA).

III-C. Sample Storage

Urine samples for each of the preparatory methods above (in triplicate) were stored in the 5 mm NMR tubes at room temperature (22°C), a refrigerator (4°C), or a deep-freeze (-80°C) for the 4 week duration of the study. Once a week samples were removed from storage and allowed to equilibrate to room temperature (roughly 1 hour) prior to NMR data acquisition.

To investigate the effects of freeze-thaw cycles on metabolites found in the urine, additional samples were prepared from the male and female subjects. Male and female urine samples were prepared as raw urine and another with 10 mM azide (see sample preparation outlined above); both were stored at -80°C and thawed twice a week (left at room temperature for 1 hour and then returned to the freezer).

III-D. NMR Analysis

All ^{1}H-NMR spectra were acquired on a 500 MHz Inova (Varian Inc, Palo Alto, CA) spectrometer equipped with a 5 mm triple-resonance probe with z-axis pulsed field gradients. One-dimensional ^{1}H-NMR spectra were collected at 25°C using a standard presaturation pulse sequence (one-dimensional, transmitter pre-saturation delay of 1 s for water suppression, followed by a 8.4 μs 90° read pulse), and a spectral width of 8000 Hz [20]. The time-domain data points were 64k, acquisition time was 4s, repetition time was 5s, four steady state scans, and the number of acquired scans was 128. The FID was apodized with an exponential window function corresponding to a line broadening of 0.5 Hz and Fourier transformed.

III-E. NMR Quantification

The methyl singlet of the buffer constituent DSS served as internal standard for chemical shifts (set to 0 ppm), and for quantification. Spectral quantification of 55 clearly identifiable metabolites were chosen for quantification using the Chenomx NMR Suite Professional software package Version 3.1 (database available at pH 7.0, Chenomx Inc., Edmonton, AB). For a list of the metabolites identified by NMR refer to Table III-1. For qualitative and quantitative analysis Chenomx NMR Suite software took a database of pure compounds (metabolites) and compared the spectral signatures to those found in the urine spectrum. Reference spectra stored in the software are fit to the urine spectra to quantify selected metabolites. The internal DSS signal was utilized as the concentration reference (0.49 mM). Work in our laboratory has demonstrated that this procedure provides absolute concentration accuracies of 90% or better[21].

Typically urine metabolites are reported as ratios with the metabolite creatinine. As a result of the changing concentration of creatinine over long-term storage in this study all metabolites were reported as absolute concentrations (mM). The metabolite urea is not listed in Table III-1. Caution must be used when drawing any conclusions based upon changes in the concentration of urea. Solvent suppression by presaturation (pulse sequence) lends to the likelihood of resonant suppression of the urea peak due to proton exchange with water. As a result urea is not included as a metabolite, but allowing for this caveat, it will be discussed later in this chapter as a potential metabolite of interest.

1,3-Dimethylurate
1,6-Anhydro-D-glucose
2-Hydroxyisobutyrate
2-Oxoglutarate
3-Hydroxybutyrate
3-Hydroxyisovalerate
3-Methylxanthine
3-Phenylpropionate
4-Aminobutyrate
4-Hydroxybutyrate
Acetate
Acetoacetate
Adenine
Alanine
Arginine
Aspartate
Benzoate
Citrate
Creatine
Creatinine
Cysteine
Dimethylamine
Ethanol
Ethanolamine
Formate
Fucose
Glucose
Glycerol
Glycine
Glycolate
Guanidoacetate
Hippurate
Histidine
Indole-3-acetate
Lactate
Malonate
Methylmalonate
Methylsuccinate
N,N-Dimethylglycine
N-Methylhistidine
O-Phosphocholine
p-Cresol
Phenylacetate
Phenylacetylglycine
Ribose
Salicylate
Sarcosine
Succinate
Taurine
Threonine
Trimethylamine
Tryptophan
Tyrosine
Uracil
Valine

Table III-1: Urine Metabolites

a. List of metabolites identified by Chenomx
b. NMR Suite Professional software v. 3.1

III-F. Statistical Analysis

Differences between variables were assessed using paired t-tests and analysis of variance (ANOVA). A p-value of < 0.05 was considered significant.

RESULTS and DISCUSSION

Studies have investigated different properties of urine; bactericidal factors[22, 23], precipitation potentials[24], the array of metabolites present[25], and even the effect of different preservatives in urine collection tubes[26, 27]. Such studies have provided a basic understanding of urine as a biofluid that can be utilized in the laboratory for a number of investigative studies concerning changes in an organisms' metabolism. However, before any differences in urine metabolites are correlated to a biological response it is important to address possible causes of changing metabolite concentrations. My study investigates the influence laboratory sample preparation and storage can have on the profile of metabolites in a normal urine sample.

III-A. Gender Differences

Throughout this study differences were noted in the metabolites that changed for male and female urine samples over the 4 week period. Of the 55 metabolites I chose to follow in female urine: acetate, benzoate, creatine, glycine, lactate, malonate, succinate, trimethylamine, and formate increased; while creatinine, urea, guanidinoacetate, hippurate and citrate decreased over the 4 weeks. Table III-2 lists the changes in female urine metabolite concentrations after a 4 week period of storage at room temperature.

Initial	Acetate	Benzoate	Citrate	Creatine	Creatinine	Formate	Glycine	Hippurate	Lactate	Malonate	Succinate	Trimethylamine	Uracil	Urea
Raw	0.00	0.00	2.74	0.30	13.1	0.22	1.49	1.69	0.21	0.07	0.02	0.07	0.18	1000
Spun	0.00	0.00	2.87	0.30	13.1	0.22	1.49	1.69	0.21	0.07	0.02	0.07	0.05	1000
Filtered	0.00	0.00	2.58	0.27	12.7	0.22	1.40	1.54	0.21	0.07	0.02	0.07	0.06	42.5
0.1mM Azide	0.00	0.00	2.66	0.30	13.1	0.22	1.49	1.69	0.21	0.07	0.02	0.07	0.08	42.7
1.0mM Azide	0.00	0.00	2.74	0.30	13.4	0.22	1.49	1.69	0.23	0.05	0.02	0.07	0.05	1000
10mM Azide	0.00	0.00	3.03	0.30	14.4	0.22	1.61	1.95	0.26	0.07	0.02	0.07	0.05	1000
Week 4														
Raw	4.66	1.02	1.00	1.38	10.1	0.76	2.42	0.06	0.75	0.15	1.03	0.57	0.26	29.9
Spun	0.82	1.08	2.47	1.27	9.26	0.49	2.08	0.09	0.79	0.09	0.13	0.06	0.06	48.5
Filtered	0.20	0.09	2.16	1.17	9.44	0.24	1.09	1.76	0.23	0.09	0.17	0.05	0.06	40.7
0.1mM Azide	0.23	1.02	2.27	1.03	9.57	0.24	1.79	0.31	0.51	0.11	0.16	0.05	0.03	40.3
1.0mM Azide	0.22	1.02	2.44	1.18	10.1	0.25	2.03	0.44	0.31	0.06	0.18	0.06	0.07	45.1
10mM Azide	0.22	0.26	2.43	1.22	10.1	0.26	1.53	1.45	0.29	0.09	0.16	0.05	0.06	47.9

Table III-2: Female Urine Preparation

a. Female urine samples prepared by different methods and stored at room temperature
b. Metabolite concentrations reported as millimolar

In the male urine: creatinine, creatine, and phenylacetylglycine were the only metabolites that changed significantly following storage at room temperature (22°C) over 4 weeks. Table III-3 lists the metabolites in the male urine that changed significantly.

Initial	Creatine	Creatinine	Phenylacetylglycine	Urea
Raw	0.48	15.1	0.77	58.2
Spun	0.48	17.8	0.77	67.2
Filtered	0.48	17.2	0.77	65.1
0.1mM Azide	0.48	16.9	0.77	64.0
1.0mM Azide	0.35	17.1	0.77	68.3
10mM Azide	0.48	16.5	0.77	63.0
Week 4				
Raw	1.71	12.2	2.77	58.2
Spun	2.26	11.9	3.05	66.6
Filtered	2.00	11.5	2.81	63.6
0.1mM Azide	2.08	12.3	2.62	66.6
1.0mM Azide	2.04	11.8	2.60	66.1
10mM Azide	2.15	10.8	2.82	70.8

Table III-3: Male Urine Preparation

a. Male urine samples prepared by different methods and stored at room temperature
b. Metabolite concentrations reported as millimolar

III-B. Urine Preparation

This study was concerned with maintaining the composition of metabolites in a normal urine sample over a period of time. The efficacy of each preparatory step was determined according to the fidelity of the metabolite profile over the 4 week period. Table III-2 shows the effect urine preparation has on the degree of metabolite change of female urine over time. Spinning the urine in a counter top centrifuge reduced the degree of change in the metabolites over the 4 weeks. Filtering the urine produced an even greater reduction in the degree of metabolite change. Increasing the concentration of the preservative sodium azide (0.1, 1, and 10 mM) had a correlated rise in efficacy of inhibition of metabolite changes over the 4 week period. Both male and female urine samples with 10 mM of sodium azide demonstrated the greatest degree of metabolite stability (Tables III-2 and 3). Across all methods of urine preparation the male urine had a smaller degree of change in metabolites when compared to the female urine.

Figure III-1 shows the change in absolute concentrations of three random metabolites (benzoate, creatine, and lactate) in female urine that was prepared raw, or following filtration, and stored at room temperature for 4 weeks. Benzoate and lactate demonstrated a dramatic increase in concentration within the first week of storage, followed by only a slight increase over the next 3 weeks for raw urine. However, following filtration benzoate and lactate did not change significantly over the 4 weeks. Both raw and filtered female urine showed a steady increase in creatine concentration over 4 weeks of storage at room temperature.

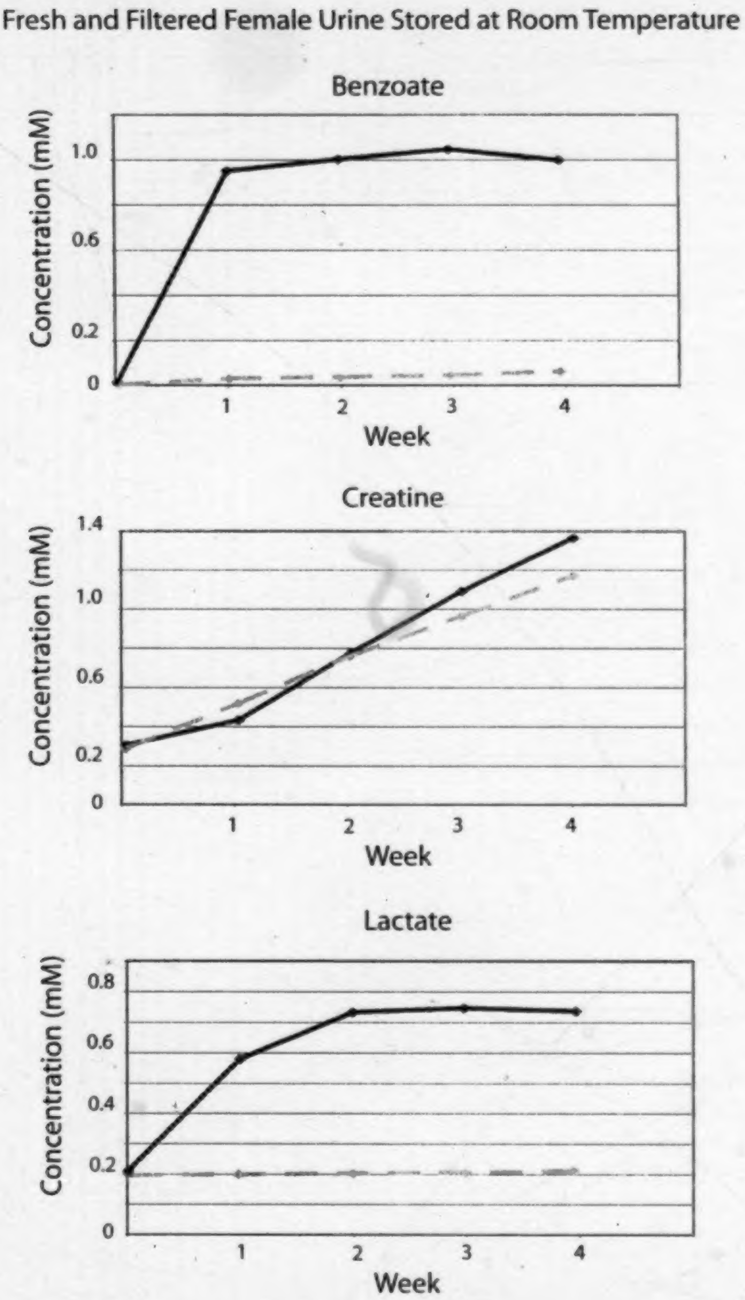


Figure III-1. **Metabolite changes in raw and filtered urine**. Absolute concentrations of three random metabolites: benzoate, creatine, and lactate, in fresh (red, solid line) and filtered (blue, dotted line) normal female urine stored at room temperature (22°C) over a 4 week period.

III-C. Urine Storage

Primarily, human urine collection for research studies takes place at a clinic or hospital setting. In some instances it can occur at the residence of the volunteer, so it is important to determine proper urine storage techniques to ensure accurate metabolite analysis at a later time. For my study urine samples were stored at room temperature (22°C), in a refrigerator (4°C), and a deep freeze (-80°C) for a 4 week period. Table III-4 demonstrates the change in raw female urine metabolite concentrations relative to the method of storage over the 4 week period. Both male and female urine samples stored at room temperature for 4 weeks had a significant change in a number of metabolites. Storage of the urine in a refrigerator produced a slight reduction in the degree of metabolite change, but storage in the deep-freeze provided urine with a metabolite profile that best reflected the original metabolite concentrations (comparing raw urine only, Table III-4). The raw urine samples that underwent repeated cycles of freeze/thaw over the 4 weeks had an intermediate degree of metabolite change when compared with raw urine stored at room temperature and in the deep-freeze (Table III-4).

	Acetate	Benzoate	Citrate	Creatine	Creatinine	Formate	Glycine	Hippurate	Lactate	Malonate	Succinate	Trimethylamine	Urea
Initial													
Room Temp	0.00	0.00	2.74	0.30	13.1	0.22	1.49	1.69	0.21	0.07	0.02	0.07	1000
Week 4													
Room Temp	4.66	1.02	1.00	1.38	10.1	0.76	2.42	0.06	0.75	0.15	1.03	0.57	29.9
Fridgerator	0.17	0.11	2.55	0.29	10.9	0.23	1.35	1.73	0.25	0.12	0.09	0.05	55.9
Deep Freeze	0.13	0.09	2.45	0.22	11.4	0.24	1.31	1.64	0.23	0.13	0.09	0.06	52.0
Freeze/Thaw	0.16	0.19	2.36	0.30	11.2	0.17	1.33	1.52	0.21	0.11	0.09	0.06	47.0

Table III-4: Female Urine Storage

a. Raw female urine samples stored under different conditions

b. Metabolite concentrations reported as millimolar

Figure III-2 shows changes in absolute concentrations of three randomly selected metabolites in raw female urine over a 4 week period of storage at either room temperature or in a deep-freeze. The metabolites citrate and glycine had a slight decrease in concentration when stored in the deep-freeze. When stored at room temperature citrate showed a significant and linear decrease in concentration, while glycine had a significant increase in concentration over the 4 weeks. The metabolite hippurate remained fairly constant when stored in the deep-freeze, but when raw female urine was stored at room temperature hippurate dropped within 1 week and then remained unchanged.

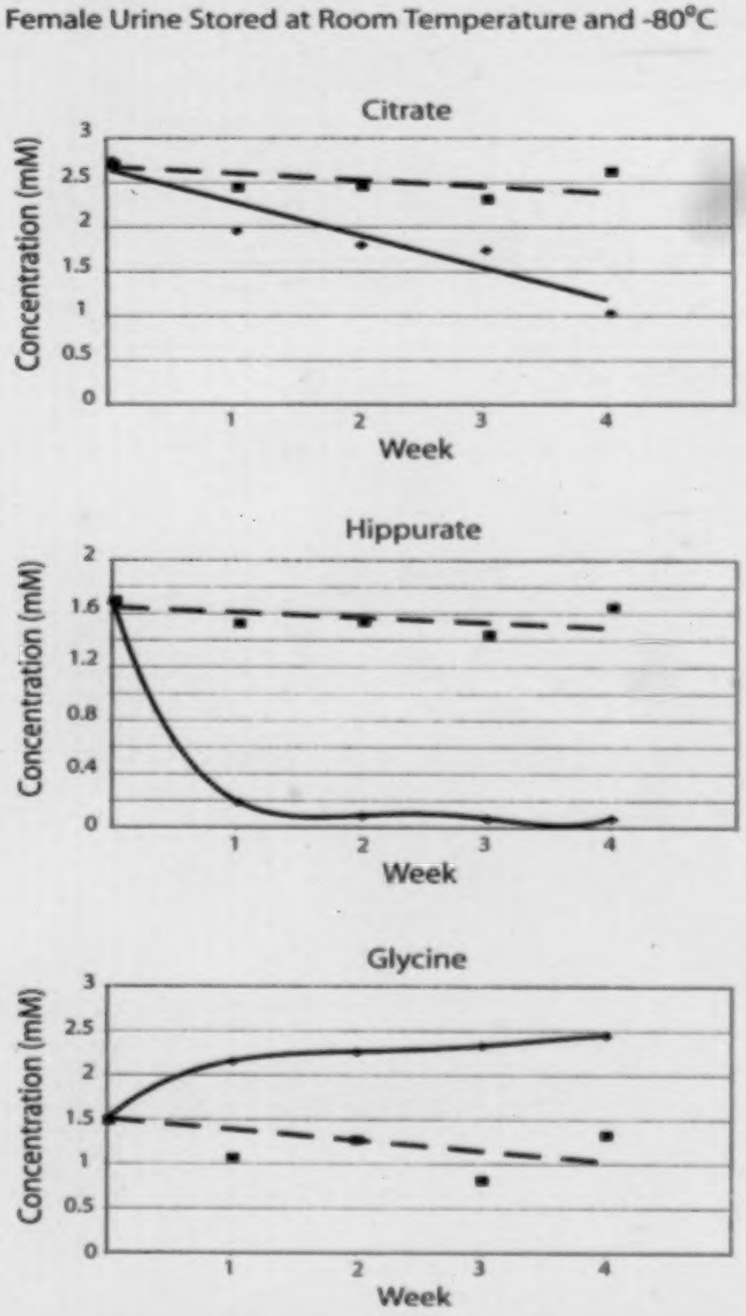


Figure III-2. **Female urine stored in a deep-freeze and at room temperature.** Concentration of citrate, hippurate, and glycine in fresh female urine stored at room temperature (red, solid line, 22°C) and in a deep-freeze (blue, dotted line, -80°C) over a 4 week period.

Similar differences in the rate and degree of metabolite change were present in female urine that was spun on the countertop centrifuge. Figure III-3 shows absolute concentration changes for three randomly chosen metabolites as identified in spun female urine and stored at either room temperature or in a deep-freeze. Benzoate and lactate showed significant increases in concentration during the first 2 weeks of storage at room temperature, with a reduction in the rate of change during the last 2 weeks of analysis. Creatine demonstrated a linear increase in concentration throughout the 4 weeks of storage at room temperature. For all three metabolites there was no significant change in absolute concentration when spun female urine was stored in the deep-freeze over 4 weeks.

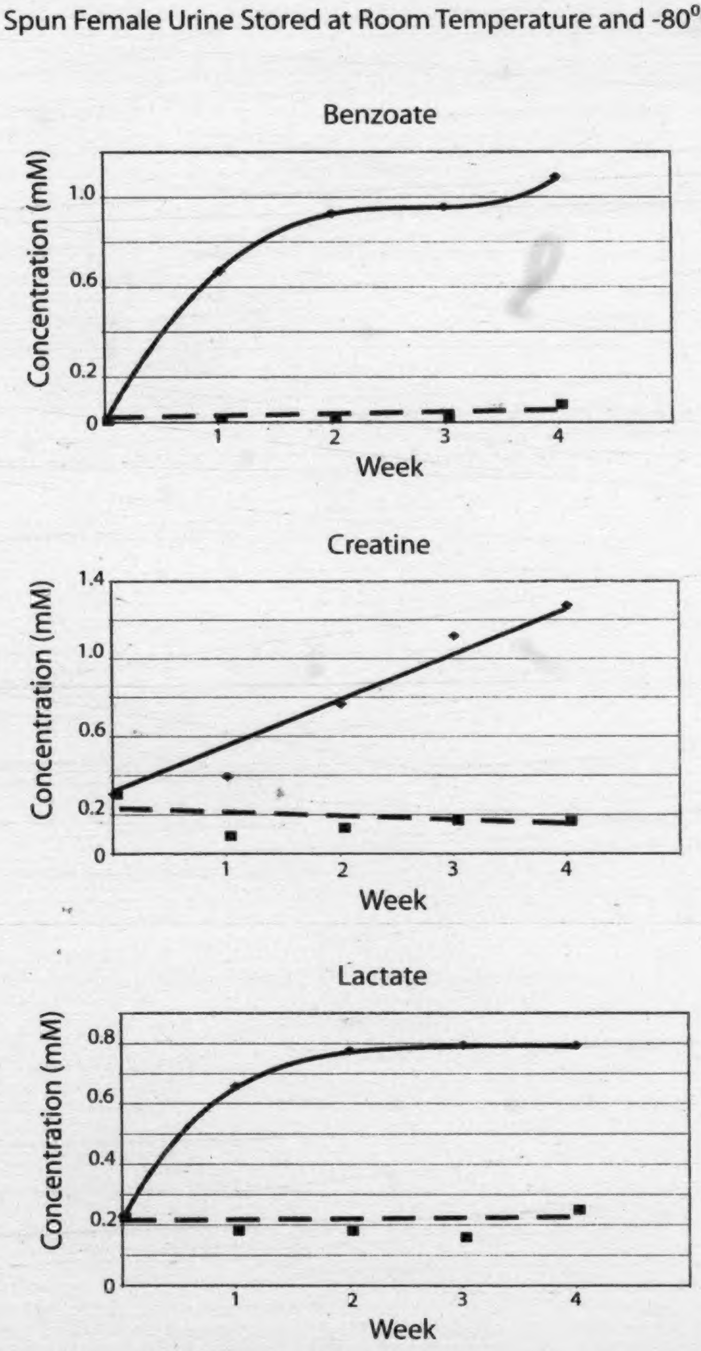


Figure III-3. **Spun urine stored in a deep-freeze and at room temperature.** Absolute concentrations for three random metabolites: benzoate, creatine, and lactate, measured in centrifuged female urine stored at room temperature (red, solid line, 22°C) and in a deep-freeze (blue, dotted line, -80°C) over a 4 week period.

III-D. Bacterial Contribution

In this study female urine benzoate concentrations increased significantly over time (Table III-2 and Figure III-3). Studies have found that when laboratory animals are moved from sterile to ambient-air environments there is an associated increase in the presence of benzoate in the urine[19, 28, 29]. This increase has been correlated with increased bacterial activity in the animals. In this study the increase in the formation of benzoate in female urine, and the reduction of benzoate formation in urine that was either filtered or possessed sodium azide suggests that bacteria in the urine may be the cause of such a change. It is known that some bacterial species are able to metabolize hippurate to benzoic acid [29]. This study found increased concentrations of benzoic acid with a correlated decrease in hippurate in the female urine; again, this is suggestive of bacterial activity in the sample (Table III-2, Figures III-2 and 3).

Over the 4 week period raw female urine demonstrated a significant increase in the concentration of the metabolite trimethylamine (Tables III-2 and 4). Previous studies have shown that trimethylamine is produced by the bacterial breakdown of dietary choline [30-32]. As well, bacteria are known to produce formic acid from the substrate citrate[33]. In the present study raw female urine showed a large increase in formic acid with a reduction in the metabolite citrate. The correlated rise of formate and fall of citrate concentrations in female urine stored over 4 weeks suggests bacterial activity in the female urine. The limiting of these absolute concentration changes by particular sample preparation (*i.e.* filtering) and storage (*i.e.* deep-freeze) techniques leads to the conclusion that such measures should be taken when using urine as a biofluid for metabolomic studies.

The concentration of creatinine fell for both the male and female urine samples, while the concentration of creatine increased. Studies have shown that creatinine in the urine can be used by bacteria and can be converted by the enzyme creatininase to creatine[34, 35]. The coordinated reduction of creatinine with an increase in creatine concentration is suggestive of bacterial activity in both male and female urine samples. Figure III-1 shows that filtering female urine had a limited reduction in the degree of change in creatine concentration over 4 weeks (stored at room temperature). It is known that creatinine is formed non-enzymatically from creatine over a period of time[36]. However, the data shows that creatinine concentrations fell while creatine concentrations grew. This may provide further evidence of the predominantly bacterial influence on metabolite levels over time as opposed to non-enzymatic reactions.

Similar to creatinine, urea is a metabolite found in urine that is rapidly broken-down by bacteria. Numerous investigations have looked at the breakdown of urea by bacteria; the formation of urinary stones from urease-producing bacteria[37], the blockage of urine-collecting systems[38], and breath tests for H.

pylori detection in gastric ulcer patients[39, 40]. My study demonstrated a reduction in the concentration of urea in normal urine over time. This reduction in urea concentration was ameliorated somewhat by centrifugation, filtration, and the addition of the preservative sodium azide when compared to raw urine. Storage in a refrigerator or deep-freeze limited the loss of urea somewhat, however, no storage method was able to halt the significant reduction in urea.

The absolute concentrations for succinate and malonate increased in female urine over the 4 week period (Tables III-2 and 4). Although sample preparation steps reduced the rise in the metabolites, there remained a significant increase in these two metabolites over long-term storage. Other studies have shown that these metabolites increase over time as a result of the activity of bacteria in urine[19, 41, 42].

The metabolite lactate also increased over time. Lactate increased in the female urine over time and once again may be the result of metabolic activity of bacteria in the urine sample. Processing of the urine and storage in the refrigerator or deep-freeze helped to reduce the rise in lactate when compared to raw urine at room temperature (Table III-4). An increase in lactate has been shown to correlate with bacterial activity in a number of pathophysiologies; such as bacteraemia[43], urinary tract infections[44], pneumonia[45], and meningitis[46, 47].

Another metabolite that increased significantly in raw female urine was acetate. Storage of urine in a refrigerator, deep-freeze, filtering, or the addition of sodium azide reduced the increase in acetate significantly (Tables III-2 and III-4). Acetate, formate, lactate, and succinate are known to rise significantly during bacterial metabolism in media where carbohydrate substrates are present[48, 49]. This study has shown that these metabolites increase over time and different methods of sample preparation and storage has an effect of reducing the amount of change in the urine metabolites. When considering the changing metabolites as a whole the data suggests that bacterial activity in the urine is changing the metabolite profile of the original sample.

The presence of bacteria in human urine is well known to the scientific and medical communities. Urinary tract infections[50-52] and the influence of gut microflora on metabolites found in the urine are all examples of the influence bacteria on detectable metabolites in human urine [53]. The metabolites that changed from the original urine samples over the 4 week period of this study are highly suggestive of bacterial contamination of the urine. This study suggests that the presence of bacteria in normal urine samples can dramatically affect the metabolic profile of the urine and appropriate measures must be taken to preserve the sample.

Differences in the degree of metabolite change between male and female urine samples over time, as well as the number of changing metabolites are possibly due to the differences in the degree of bacterial contamination. It is possible that the urological differences between male and female urinary tracts allows for greater bacterial culture and urine contamination in female patients[54].

My study focused primarily upon sample preparation and storage of normal urine samples. Urine samples were collected from a male and female volunteer. Due to biological variance observed in human urine another extended study would be beneficial. A number of individuals could be sampled and the effect of biological variability on the degree of metabolite change in stored urine over time may be determined.

CONCLUSION

Urine is an attractive biofluid for metabolomic and metabonomic investigations due to the ease of sample collection and the rich metabolite nature of the fluid. However, as investigators identify and follow changes in urine metabolite concentrations it becomes increasingly important to define the origin of the observed differences. This study has shown that appropriate storage and preparatory measures must be taken immediately following urine collection to that ensure the metabolite composition of the sample is maintained throughout the collection and analytical process. The bacteriostatic preservative sodium azide should be added to all urine samples, the samples should be stored in a deep-freeze, and cycles of sample freeze-thaw should be avoided whenever possible. The filtration of urine provided additional preservation of original metabolite concentrations, likely due to the removal of bacterial contamination, but caution is advised in some cases due to the possibility of losing larger metabolites, such as some proteins, during the filtration process. Investigators must take appropriate measures to remove or inhibit bacterial influence upon the metabolic profile of the urine sample. Standard procedures for urine preparation and storage and the continued monitoring of any metabolite changes will ensure sample fidelity for future metabolomic investigations.

LITERATURE CITED

1. Weckwerth, W. and K. Morgenthal, Metabolomics: from pattern recognition to biological interpretation. Drug Discov Today, 2005. 10(22); 1551-8.
2. Lindon, J.C., et al., Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. Biomarkers, 2004. 9(1); 1-31.
3. Bollard, M.E., et al., NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. NMR Biomed, 2005. 18(3); 143-62.
4. Moolenaar SH, U.F. Engelke, and Wevers RA, Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism. Ann Clin Biochem, 2003. 40(Pt1); 16-24.
5. Dunn, W.B., N.J. Bailey, and H.E. Johnson, Measuring the metabolome: current analytical technologies. Analyst, 2005. 130(5); 606-25.
6. Chadha, V., U. Garg, and U.S. Alon, Measurement of urinary concentration: a critical appraisal of methodologies. Pediatr Nephrol, 2001. 16(4); 374-82.
7. Griffiths, L., Assay by nuclear magnetic resonance spectroscopy:quantification limits. Analyst, 1998. 123; 1061-1068.
8. Coen, M., et al., Proton nuclear magnetic resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis. Clin Infect Dis, 2005. 41(11); 1582-90.
9. Sabatine, M.S., et al., Metabolomic identification of novel biomarkers of myocardial ischemia. Circulation, 2005. 112(25); 3868-75.
10. Brindle, J.T., et al., Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med, 2002. 8(12); 1439-44.

11. Zuppi, C., et al., 1H NMR spectra of normal urines: reference ranges of the major metabolites. Clin Chim Acta, 1997. 265(1); 85-97.

12. Symanski, E. and N.M. Greeson, Assessment of variability in biomonitoring data using a large database of biological measures of exposure. AIHA J (Fairfax, Va), 2002. 63(4); 390-401.

13. Guneral, F. and C. Bachmann, Age-related reference values for urinary organic acids in a healthy Turkish pediatric population. Clin Chem, 1994. 40(6); 862-6.

14. Dyer, A., et al., Urinary biochemical markers of dietary intake in the INTERSALT study. Am J Clin Nutr, 1997. 65(4 Suppl); 1246S-1253S.

15. Lenz, E.M., et al., A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. J Pharm Biomed Anal, 2003. 33(5); 1103-15.

16. Lenz, E.M., et al., Metabonomics, dietary influences and cultural differences: a 1H NMR-based study of urine samples obtained from healthy British and Swedish subjects. J Pharm Biomed Anal, 2004. 36(4); 841-9.

17. Holmes, E., et al., Chemometric models for toxicity classification based on NMR spectra of biofluids. Chem Res Toxicol, 2000. 13(6); 471-8.

18. Tate, A.R., S.J. Damment, and J.C. Lindon, Investigation of the metabolite variation in control rat urine using (1)H NMR spectroscopy. Anal Biochem, 2001. 291(1); 17-26.

19. Nicholls, A.W., R.J. Mortishire-Smith, and J.K. Nicholson, NMR spectroscopic-based metabonomic studies of urinary metabolite variation in acclimatizing germ-free rats. Chem Res Toxicol, 2003. 16(11); 1395-404.

20. Hoult, D.I., Solvent peak saturation with single phase and quadrature Fourier transformation. Journal of Magnetic Resonance, 1976. 21; 337 - 347.

21. Saude, E.J. and B.D. Sykes, Urine stability for metabolomic studies: effects of preparation and storage. Metabolomics, 2006. In Press.

22. Griffith, D.P., D.M. Musher, and C. Itin, Urease. The primary cause of infection-induced urinary stones. Invest Urol, 1976. 13(5); 346-50.

23. Kaye, D., Antibacterial activity of human urine. J Clin Invest, 1968. 47(10); 2374-90.

24. Udert, K.M., T.A. Larsen, and W. Gujer, Estimating the precipitation potential in urine-collecting systems. Water Res, 2003. 37(11); 2667-77.

25. Armstrong, M.D., K.N. Shaw, and P.E. Wall, The phenolic acids of human urine; paper chromatography of phenolic acids. J Biol Chem, 1956. 218(1); 293-303.

26. Allen, T.A., R.L. Jones, and J. Purvance, Microbiologic evaluation of canine urine: direct microscopic examination and preservation of specimen quality for culture. J Am Vet Med Assoc, 1987. 190(10); 1289-91.

27. Nickander, K.K., C.J. Shanholtzer, and L.R. Peterson, Urine culture transport tubes: effect of sample volume on bacterial toxicity of the preservative. J Clin Microbiol, 1982. 15(4); 593-5.

28. Goodwin, B.L., C.R. Ruthven, and M. Sandler, Gut flora and the origin of some urinary aromatic phenolic compounds. Biochem Pharmacol, 1994. 47(12); 2294-7.

29. Hansen, S., et al., Urinary bacteria: potential source of some organic acidurias. Clin Chim Acta, 1972. 39(1); 71-4.

30. Zeisel, S.H., et al., Conversion of dietary choline to trimethylamine and dimethylamine in rats: dose-response relationship. J Nutr, 1989. 119(5); 800-4.

31. Brewster, M.A. and H. Schedewie, Trimethylaminuria. Ann Clin Lab Sci, 1983. 13(1); 20-4.

32. Zeisel, S.H., K.A. DaCosta, and J.G. Fox, Endogenous formation of dimethylamine. Biochem J, 1985. 232(2); 403-8.

33. Schiwara, H.W., H. Siegel, and A. Goebel, Increase and decrease in formic acid concentration in urine samples stored at room temperature. Eur J Clin Chem Clin Biochem, 1992. 30(2); 75-9.

34. Jones, J.D. and P.C. Burnett, Implication of creatinine and gut flora in the uremic syndrome: induction of "creatininase" in colon contents of the rat by dietary creatinine. Clin Chem, 1972. 18(3); 280-4.

35. Cattell, W.R., et al., Creatinine content of urine and bacterial growth. Br Med J, 1969. 3(663); 175-6.

36. Wyss, M. and R. Kaddurah-Daouk, Creatine and creatinine metabolism. Physiol Rev, 2000. 80(3); 1107-213.

37. Kaya, S., et al., Role of genital mycoplasmata and other bacteria in urolithiasis. Scand J Infect Dis, 2003. 35(5); 315-7.

38. Udert, K.M., et al., Urea hydrolysis and precipitation dynamics in a urine-collecting system. Water Res, 2003. 37(11); 2571-82.

39. Pathak, C.M., D.K. Bhasin, and K.L. Khanduja, Urea breath test for Helicobacter pylori detection: present status. Trop Gastroenterol, 2004. 25(4); 156-61.

40. Pathak, C.M., et al., 14C-urea breath test as a 'gold standard' for detection of Helicobacter pylori infection. Med Sci Monit, 2004. 10(8); LE14-5.

41. Brenner, D.J., et al., Enterobacter asburiae sp. nov., a new species found in clinical specimens, and reassignment of Erwinia dissolvens and Erwinia nimipressuralis to the genus Enterobacter as Enterobacter dissolvens comb. nov. and Enterobacter nimipressuralis comb. nov. J Clin Microbiol, 1986. 23(6); 1114-20.

42. Stenina, M.A., et al., Tissue hypoxia and intestinal dysbiosis in children with tuberculosis. Bull Exp Biol Med, 2003. 135(2); 178-80.

43. Bar-Meir, M., et al., Non-Typhi Salmonella gastroenteritis in children presenting to the emergency department: characteristics of patients with associated bacteraemia. Clin Microbiol Infect, 2005. 11(8); 651-5.

44. Tal, S., et al., Profile and prognosis of febrile elderly patients with bacteremic urinary tract infection. J Infect, 2005. 50(4); 296-305.

45. Utine, G.E., et al., Childhood parapneumonic effusions: biochemical and inflammatory markers. Chest, 2005. 128(3); 1436-41.

46. Nazifi, S., A. Rezakhani, and M. Badran, Evaluation of hematological, serum biochemical and cerebrospinal fluid parameters in experimental bacterial meningitis in the calf. Zentralbl Veterinarmed A, 1997. 44(1); 55-63.

47. Lutsar, I., et al., Five days of antibacterial therapy for bacterial meningitis in children? Infection, 1995. 23(2); 113-8.

48. Macfarlane, S. and G.T. Macfarlane, Regulation of short-chain fatty acid production. Proc Nutr Soc, 2003. 62(1); 67-72.

49. D'Argenio, G. and G. Mazzacca, Short-chain fatty acid in the human colon. Relation to inflammatory bowel diseases and colon cancer. Adv Exp Med Biol, 1999. 472; 149-58.

50. Funfstuck, R., et al., The influence of selected urinary constituents on the adhesion process of Escherichia coli to human uroepithelial cells. Clin Nephrol, 1987. 28(5); 244-9.

51. Peddie, B.A., S.T. Chambers, and M. Lever, Is the ability of urinary tract pathogens to accumulate glycine betaine a factor in the virulence of pathogenic strains? J Lab Clin Med, 1996. 128(4); 417-22.

52. Chambers, S.T., et al., Inhibitors of bacterial growth in urine: what is the role of betaines? Int J Antimicrob Agents, 1999. 11(3-4); 293-6.

53. Bain, M.D., et al., Contribution of gut bacterial metabolism to human metabolic disease. Lancet, 1988. 1(8594); 1078-9.

54. Shopova, E., A. Nikolov, and A. Dimitriv, [Link between the state of vaginal flora and the development of uroinfection during pregnancy]. Akush Ginekol (Sofiia), 2005. 44(1); 38-9.

CHAPTER IV

Variation of Metabolites in Normal Human Urine

OVERVIEW

Urine is often sampled from patients participating in clinical and metabolomic studies. Biological homeostasis is known to occur in humans, but little is known about the variability of metabolites found in urine. Scientific studies rely upon baseline measurements in order to identify and quantify differences and changes. In clinical and metabolomic studies the baseline that is typically chosen is a normal or a control population. Scientists use the comparison of a studied or diseased group with the normal group to highlight the significant biological changes or pathophysiological processes involved. However, very little is known regarding the metabolite variability of human urine. This must be studied before normal human urine is used as a baseline for future clinical and metabolomic investigations. *The study discussed in this chapter investigates the variability of selected urine metabolites in a group of healthy men and women over a period of 30 days*. To monitor individual variation, six women from the normal population were randomly selected and followed for 30 days. To determine the influence of extraneous environmental factors, urine was collected from 25 guinea pigs with similar genetics, diet, and living environment. This study identifies the possible variation of urine metabolites. In addition, this study also addressed a philosophical question "What is normal?" and initiated future discussions regarding the reconciliation of homeostasis with urine variability.

INTRODUCTION

The human body is composed of a complex and dynamic array of interacting metabolic pathways. These complex chemical reactions are generally kept in balance in a healthy individual; otherwise the slightest metabolic deviation can produce devastating results[1, 2]. It is believed that tight control over the metabolic profile of an individual, referred to as homeostasis, is required to maintain health.

Whether a primary cause, or a secondary indicator, metabolite changes have long been observed in diseased individuals[3-7]. Metabolomic studies attempt to detect, monitor, and in some cases predict the metabolic response to an insult or disease state[8]. These studies include investigations of single-celled organisms, plants, animals, tissues, organs, and human subjects[9-15]. Some metabolomic studies have attempted to understand specific metabolic pathways and processes, while others have attempted to diagnose a particular disease[1, 2, 16-20]. Therefore, such metabolic conclusions rely heavily upon having a well-defined healthy group with well-characterized metabolite homeostasis.

High-resolution NMR (most often ^{1}H-NMR) has emerged as a promising non-invasive technique for metabolomic studies due to its ability to simultaneously detect a large number of compounds in a rapid and high-throughput manner that requires little sample manipulation. The lack of invasive mechanical or chemical sample modification for NMR is important because it allows the samples to be analyzed, stored, and re-analyzed if methods improve at a later date. In addition, the data collected (spectra) may be analyzed by several techniques, including spectral integration, principal components analysis (PCA), partial least squares (PLS), and neural networks [21-23].

Currently, there is limited information available regarding normal metabolite concentrations and degrees of variance found in urine [24-26]. Due to limitations in laboratory resources most studies have tested samples from 10-30 patients, with a control or normal group of at least similar size. This limitation in the number of samples is compounded in human studies where subject compliance may be poor[27]. Although reference to Geigy scientific tables is common, the information is based upon limited research with similarly small sample sizes[26].

Most human or clinical studies recruit healthy volunteers based upon the absence of medical co-morbidities. The metabolite measurements of the normal cohort are then pooled and used for comparison with test subjects (usually those with a defined disease, *e.g.* diabetes). A certain degree of homeostasis is assumed to occur in the excretion of metabolites in normal urine; however, it is extremely important to define the inter- and intra-individual metabolite variance within the control group before conclusions are made regarding metabolite changes in diseased individuals. The definition of the normal metabolic profile is fundamental in human studies where additional variance is introduced through a number of uncontrolled factors (*e.g.* genetics, ethnicity, stress, exercise level, and diet)[6, 28]. The primary goal of this study was to investigate the degree of inter- and intra- subject metabolite variance in a group of normal healthy human subjects using ^{1}H-NMR analysis of urine to provide a normal human urine metabolic baseline necessary for future metabolomic and clinical studies. For comparison, 25 guinea pigs of the same strain, gender, diet, and environment were analyzed to investigate the influence of extraneous factors on urine metabolite excretion often observed in human clinical investigations. This human and animal model normal study identified urine metabolite variability and inter- and intra-subject variability, and possible causes of the variation.

EXPERIMENTAL PROCEDURES

IV-A. Human Control Sample Collection

Informed consent was obtained from healthy normal volunteers (30 male, 30 female) in accordance with guidelines established by the University of Alberta Health Research Ethics Board. The study attempted to sample a representative cohort of normal or 'relatively' healthy control individuals. Each volunteer was accepted based upon a physician's opinion that no chronic or major illness was present. For the normal population, each individual collected two daily mid-stream urine samples, one after waking (early AM), and another in the midafternoon. Sample collection continued for 30 days, which provided 3600 urine samples.

IV-B. Human Sample Handling

The urine samples were treated with sodium azide (Sigma Aldrich) to a final concentration of 2.5 mM immediately following collection and stored at -10°C for periods less than one week. Samples were catalogued and stored at the Canadian National High Field NMR Centre (NANUC) in a deep-freeze (-80°C). To obtain a general understanding of inter- individual metabolite variation within a 'normal' group 1 day of morning urine was randomly selected, for a total of 57 samples (30 male and 27 female). To determine intra- individual variation 6 female subjects were randomly chosen and their morning urine samples were analyzed over the 30 day period. As a result of subject compliance to regular sample collection the female subjects (randomly numbered 1 – 6) had differing number of total samples over the 30 days (21, 24, 29, 30, 29, and 29 samples respectively). To investigate metabolite differences as a result of gender one day was randomly chosen and the morning and afternoon urine samples from the female and male subjects were analyzed (total of 60 male, and 55 female samples from one day). Differences in urine metabolites due to time of sampling was investigated by using the same day selected for the 'gender analysis', and comparing the morning and afternoon urine samples from the male and female subjects (30 male samples from both morning and afternoon collections from one day; 27 female morning and 28 female afternoon samples from one day).

IV-C. Guinea Pig Control Sample Collection

Urine samples were taken from 25 female Dunkin-Hartley guinea pigs (pathogen-free, 180–450 g, Charles River Laboratories). All animals were shipped in filtered crates and kept in high-efficiency, particulate-filtered air. All animals were fed a standard guinea pig diet (Prolab; Agway) including water *ad libitum*. Animals were handled in accordance with the Canadian Council on Animal Care guidelines. The Health Sciences Animal Policy and Welfare Committee, University of Alberta approved the ethics for the use of the animals. The guinea pigs were anesthetized intraperitoneally with urethane (1.5 g/kg). Adequate depth of anesthesia was assessed via the pedal reflex. Urine samples were collected by trans-abdominal cystocentesis using a sterile 21-gauge needle. The needle was inserted at a 45° angle, on the midline, midway between the umbilicus and brim of the pelvis, while creating negative pressure by pulling back on the plunger of the syringe. If urine was not obtained with the first puncture, two additional punctures were attempted from 0.2 to 1.0 centimeters cranial or caudal to the initial puncture site. The needle was changed before making each attempt to avoid contamination. Urine samples (1.0-2.0 ml) were stored at -80°C for later ^{1}H-NMR analysis.

IV-D. Sample Preparation

Both human and guinea pig urine samples were thawed in a biosafety fume hood and a 630 μl aliquot was removed and placed in a 1.5 ml Eppendorff tube followed by the addition of 70 μl of a reference buffer solution ((4.9 mM DSS (disodium-2, 2-dimethyl 2-silapentane-5-sulphonate) and 100 mM imidazole in D_2O) Sigma-Aldrich, Mississauga, ON). Each sample was brought to a pH of 7.0 +/- 0.1 using HCl and NaOH. An aliquot of 600 μl was taken and transferred to a standard 5 mm glass NMR tube (Wilmad, NJ, USA).

IV-E. NMR Analysis

All ^{1}H-NMR spectra were acquired on a 600 MHz Inova (Varian Inc, Palo Alto, CA.) spectrometer equipped with a 5 mm triple-resonance probe with z-axis gradient coil. One-dimensional ^{1}H-NMR spectra were collected at 25°C with a tnnoesy pulse sequence (one-dimensional, three pulse NOESY, with a transmitter pre-saturation delay of 900 ms for water suppression during the pre-acquisition delay and 100 ms mixing time), and a spectral width of 7200 Hz. The time-domain data points were 64k, acquisition time was 4s, the 90° pulse was 6.8μs, repetition time was 5s, steady state scans were four, and the number of acquired scans was 32. The FID was apodized with an exponential window function corresponding to a line broadening of 0.5 Hz, zero-filled to 128k, and Fourier transformed [29].

The methyl singlet of the buffer constituent DSS served as internal standard for chemical shifts (set to 0 ppm), and quantification. Spectral identification and quantification of 24 clearly identifiable metabolites was performed using the Chenomx NMR Suite Professional software package Version 3.1 (database available at pH 7.0, Chenomx Inc., Edmonton, AB). Figure IV-1 shows the expanded aromatic (7.2 – 8.3 ppm) and aliphatic regions (3.3 – 4.7 ppm) of a 1D ^{1}H-NMR spectrum of a typical human urine sample together with the spectral regions and resonant peaks of a reference sample of pure hippuric acid. For qualitative and quantitative analysis Chenomx NMR Suite software takes a database of pure compounds (metabolites) and compares the spectral signatures to those found in the urine spectrum. A least-squares fit of the reference spectra to the urine spectra was used to quantify selected metabolites. The internal DSS signal was utilized as the concentration reference (0.49 mM). The software includes an appropriate correction for urine dilution due to the addition of buffer. Work in our laboratory has demonstrated that this procedure provides absolute concentration accuracies in excess of 90%[30].

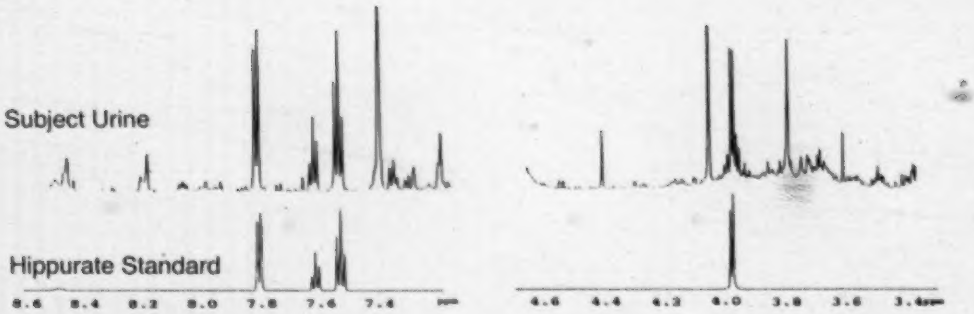


Figure IV-1. **Spectral analysis by Chenomx software**. Expanded aromatic (6.7-8.7 ppm) and aliphatic (3.3-4.7 ppm) regions of a 600 MHz 1D ^{1}H-NMR spectrum of healthy human urine (top). The bottom trace shows the reference spectrum of hippurate used by the Chenomx NMR Suite software for qualitative identification and quantitative determination for the metabolite hippurate.

Metabolite concentrations were expressed as absolute values (Figure IV-2), and as ratios relative to creatinine to correct for dilution assuming a constant rate of creatinine excretion for each urine sample[31]. Following normalization of metabolite concentrations a large degree of metabolite variability remained. Scaling the normalized metabolite concentrations according to the approximate population average concentration of creatinine (for the human samples: Eq. 1 [mM metabolite/mM creatinine] x 10, and for the guinea pig samples: Eq. 2 [mM metabolite/mM creatinine] x 3), returned metabolite values of a similar scale to the raw metabolite concentrations. This allowed for graphical comparison to determine the remaining degree of metabolite variability. Scaling does not alter the statistical metabolite variability (*i.e.* coefficients of variation remain constant); however, it does allow for graphical overlay of raw and normalized metabolite concentrations to determine if the degree of metabolite variability is reduced following dilution correction with creatinine.

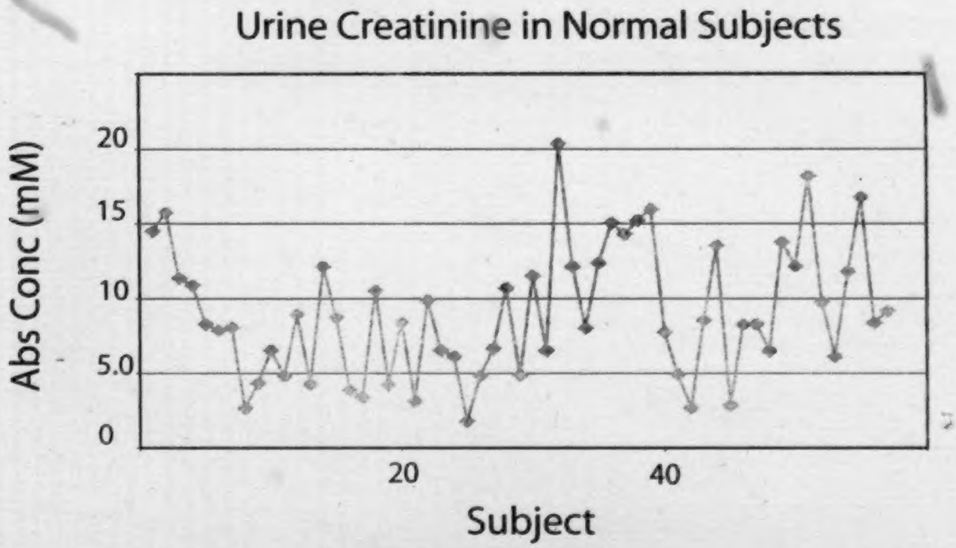


Figure IV-2. **Creatinine variability in a normal human population**. Absolute creatinine concentrations of 57 healthy subject urine samples from a single day (30 male, 27 female), and the morning sampling time (9.03 ± 4.4 mM). This sampling was randomly selected from a large dataset of serial samples collected from our normal population of 30 men and 30 women.

IV-G. Statistical Analysis

Metabolite concentration and normalized results are expressed as means, ranges, and coefficients of variation ((standard deviation/mean)*100)). Differences between variables and sub-groups were investigated using Box-Cox transformation (to ensure normality and equality of variance) of normalized metabolite values followed by one-way analysis of variance (ANOVA). A p-value of < 0.05 was considered to indicate statistical significance.

RESULTS

IV-A. Human Demographics

Urine samples were collected from 30 men with a mean age of 39.6 (± 14, range: 20 - 74), and 30 women with a mean age of 41.3 (± 16, range: 18 - 68).

IV-B. Metabolite Identification

The 24 metabolites chosen for this study are all major metabolites found in urine and were unequivocally identified and quantified with concentrations ranging from 0.01 to 10.00 mM. The mean metabolite concentrations, standard deviation, range, and coefficient of variation for the normal population are listed in Table IV-1. The variability of metabolites in the urine were analyzed as group means, coefficients of variance, and in terms of intra-subject variance over time.

Metabolite	Absolute (mM)		Coeff Var	Normalized		Coeff Var
	mean ± SD	min/max		mean ± SD	min/max	
2-Hydroxyisobutyrate	0.05 ± 0.03	(0.01/0.17)	63.83	0.06 ± 0.02	(0.02 - 0.15)	36.36
2-Oxoisocaproate	0.02 ± 0.01	(0.00/0.07)	47.62	0.03 ± 0.02	(0.00 - 0.08)	80.00
4-Aminohippurate	0.02 ± 0.02	(0.00/0.06)	100.0	0.02 ± 0.02	(0.00 - 0.12)	83.33
4-Hydroxyphenylacetate	0.11 ± 0.05	(0.02/0.24)	47.62	0.12 ± 0.05	(0.06 - 0.29)	40.32
Alanine	0.24 ± 0.16	(0.04/0.80)	66.67	0.28 ± 0.12	(0.07 - 0.73)	43.48
Aspartate	0.14 ± 0.07	(0.00/0.34)	51.85	0.17 ± 0.11	(0.00 - 0.53)	63.58
Citrate	2.54 ± 1.81	(0.05/9.50)	46.55	3.05 ± 1.98	(0.12 - 10.9)	64.83
Creatine	0.55 ± 0.49	(0.00/2.05)	88.77	0.66 ± 0.55	(0.00 - 2.54)	83.46
Creatinine	9.03 ± 4.37	(1.78/20.35)	48.37	10.0 ± 0.0	(0.00 - 10.0)	
Cytidine	0.03 ± 0.08	(0.00/0.40)	235.3	0.05 ± 0.11	(0.00 - 0.49)	239.1
Formate	0.33 ± 0.48	(0.03/3.47)	14.46	0.41 ± 0.50	(0.07 - 3.10)	121.1
Glutamate	0.14 ± 0.11	(0.00/0.59)	81.48	0.16 ± 0.12	(0.00 - 0.70)	75.95
Hippurate	1.83 ± 1.24	(0.26/5.72)	67.95	2.28 ± 1.43	(0.27 - 6.87)	62.86
Histidine	0.49 ± 0.41	(0.01/1.35)	84.36	0.53 ± 0.36	(0.02 - 1.45)	68.44
Hypoxanthine	0.04 ± 0.03	(0.00/0.15)	78.95	0.04 ± 0.03	(0.00 - 0.16)	69.77
Lactate	0.38 ± 1.28	(0.03/9.79)	334.2	0.40 ± 1.08	(0.07 - 8.28)	270.0
N-Methylhistidine	0.24 ± 0.19	(0.00/0.83)	79.83	0.25 ± 0.13	(0.00 - 0.60)	52.00
Phenylalanine	0.12 ± 0.07	(0.01/0.33)	59.32	0.14 ± 0.08	(0.04 - 0.40)	57.97
Salicylurate	0.06 ± 0.07	(0.00/0.31)	125.0	0.07 ± 0.11	(0.00 - 0.48)	148.6
Trans-Aconitate	0.04 ± 0.04	(0.00/0.21)	90.91	0.05 ± 0.04	(0.00 - 0.20)	81.63
Threonine	0.08 ± 0.06	(0.00/0.32)	80.00	0.10 ± 0.07	(0.00 - 0.32)	72.16
Tryptophan	0.11 ± 0.07	(0.02/0.28)	63.06	0.13 ± 0.07	(0.03 - 0.33)	53.44
Tyrosine	0.10 ± 0.07	(0.01/0.29)	69.31	0.13 ± 0.14	(0.02 - 1.09)	109.4
Uridine	0.00 ± 0.01	(0.00/0.06)	250.0	0.01 ± 0.01	(0.00 - 0.06)	100.0

Table IV-1. Urine metabolites analyzed

a. 24 urine metabolites were quantified for variability in normal subjects

b. Concentrations reported as millimolar; 30 male and 27 female samples

c. Creatinine was used for urine normalization (Eq. [1])

d. Buffer constituents DSS and imidazole were 0.49 and 10.0 mM respectively

e. Coefficient of variation 100*(SD/Mean)

IV-C. Normalization to Creatinine

Absolute urine metabolite concentrations are affected by urine volume, which can be correlated to the amount of liquid taken in by the diet. The compound creatinine is easily identified in the urine, is correlated to overall body size (related to muscle mass), and in present literature is said to have little metabolic or excretory variance in healthy individuals. As a result, creatinine is often used to compensate for variations in urine volume, and metabolite concentrations are commonly expressed as ratios relative to creatinine [24]. The variation of creatinine concentrations in 57 urine samples (30 male and 27 female, morning samples) from the cohort of normal subjects is shown in Figure IV-2. The mean and standard deviation (9.03 +/- 4.37 mM) demonstrates that there is substantial variation of the absolute creatinine concentration in the normal adult cohort.

IV-D. Inter-individual Variability

To determine the effect of urine volume on metabolite concentration variance and the proper method of reporting metabolite concentrations, two clearly identifiable metabolites (hippurate and citrate) were chosen from the set of 24 metabolites. Absolute hippurate concentrations for normal subjects are shown graphically in Figure IV-3 (a), (mean: 1.83 ± 1.24 mM). The normalized hippurate concentrations following individual standardization to creatinine as 10 mM are shown in Figure IV-3 (b), (mean: 2.28 ± 1.43).

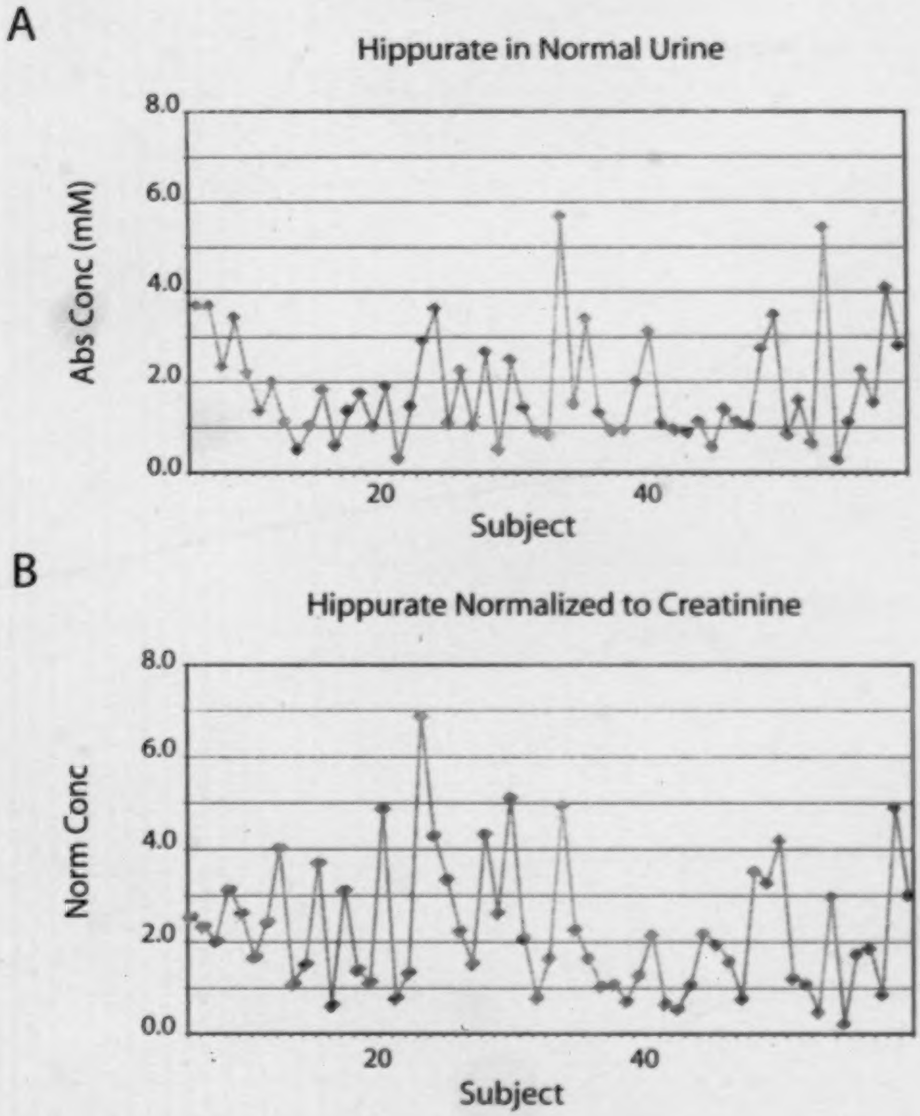


Figure IV-3. **Human urine hippurate concentration normalization**. Urine hippurate concentrations for the normal population (57 urine samples from a single day, morning sampling time, 30 male and 27 female) prior to (1.83 ±1.24 mM, A), and following normalization to creatinine (see Eq. 1., 2.28 ± 1.43, B).

This analysis was repeated for the metabolite citrate (Figure IV-4), whereby means for the absolute and scaled concentrations were 2.54 ± 1.81 mM, and 3.05 ± 1.98, respectively. The analysis of all 24 chosen metabolites revealed similarly large standard deviations and coefficients of variation for both absolute and scaled concentrations within the adult normal cohort (Table IV-1).

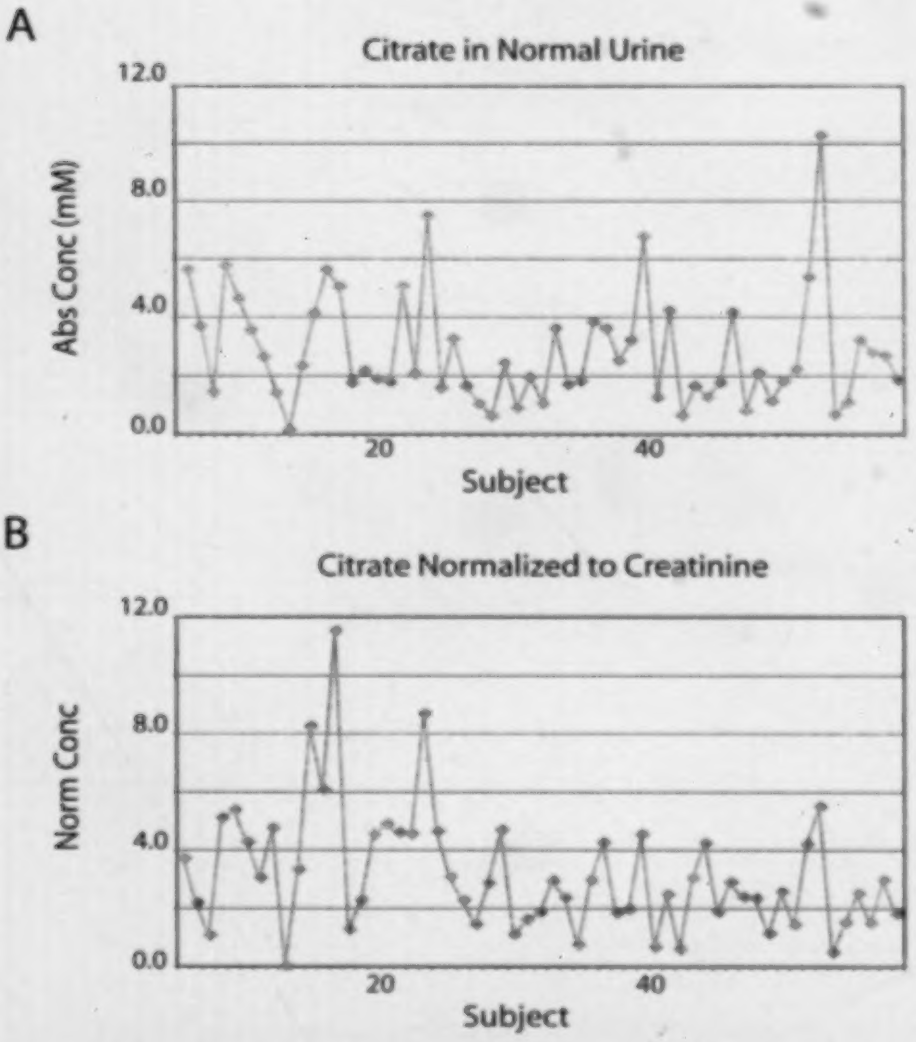


Figure IV-4. **Human urine citrate concentration normalization**. Citrate levels in the urine for the normal population (57 urine samples from a single day, morning sampling time, 30 male and 27 female) prior to (2.54 ± 1.81 mM, A), and following normalization to creatinine (3.05 ± 1.98, B).

IV-E. Gender

Within the 30 day collection period a single day was randomly selected for analysis of gender differences in urinary metabolite secretion. Morning and afternoon urine samples were included for the male and female subjects (total of 60 samples for the male, and 55 samples for the female). Following Box-Cox normalization of the concentrations for the 24 metabolites identified in the normal urine, the following twelve metabolites were found to be statistically different ($p<0.05$) for male vs. female subjects: 2-hydroxyisobutyrate, 4-aminohippurate, aspartate, citrate, creatine, formate, hippurate, histidine, lactate, threonine, trans-aconitate, and N-methylhistidine.

IV-F. Sampling Time

The same random day of urine collection selected for the analysis of gender differences was used to determine differences in urine metabolite concentrations as a result of urine collection time. Male and female morning urine samples for the day (57 samples) were included as the 'morning urine samples', as were the male and female afternoon urine collections (58 urine samples). Following Box-Cox normalization of the metabolite concentrations, the following nine metabolites were found to be statistically different ($p<0.05$): 4-aminohippurate, aspartate, creatine, formate, glutamate, phenylalanine, salicylurate, tryptophan, and trans-aconitate.

IV-G. Intra-individual Variability

From the initial 57 subjects in our randomly selected normal population, six female subjects were randomly chosen and 10 unequivocally identifiable metabolites were followed over a 4 week period using morning voids only. Metabolite concentrations were normalized to creatinine levels according to Eq. [1] (see Methods section). Longitudinal values, the mean, and the coefficient of variance for six randomly selected normalized metabolites are shown in Table IV-2. Morning citrate concentrations for six normal female subjects over the 4 week period are shown in Figure IV-5. The data demonstrates that normalized citrate concentrations for each individual remain relatively stable for periods of time; however, during other periods of time vary by over one-to-two times the female cohort standard deviation. Morning tyrosine normalized concentrations for the same six female subjects, over the same 4 week period, are shown in Figure IV-6. In contrast, individual tyrosine variation is ≤ 1 cohort standard deviation, except for one aberrant spike for female #2.

Subject	Alanine	Coeff Var	Citrate	Coeff Var	Creatine	Coeff Var	Hippurate	Coeff Var	Lactate	Coeff Var	Tyrosine	Coeff Var
Female 1	0.17 ± 0.05	30.30	3.89 ± 0.92	23.66	0.71 ± 0.39	55.16	2.44 ± 1.51	61.96	0.16 ± 0.05	30.86	0.04 ± 0.01	27.78
Female 2	0.16 ± 0.07	44.30	2.07 ± 0.69	33.40	1.09 ± 0.97	89.32	1.41 ± 0.53	24.75	0.18 ± 0.07	38.04	0.05 ± 0.06	122.4
Female 3	0.22 ± 0.07	31.82	4.34 ± 1.77	40.83	0.96 ± 0.73	75.96	1.11 ± 0.87	78.38	0.22 ± 0.25	115.7	0.04 ± 0.01	23.81
Female 4	0.24 ± 0.05	20.83	6.05 ± 1.55	25.62	0.54 ± 0.66	121.8	2.07 ± 0.90	43.39	0.18 ± 0.06	32.61	0.07 ± 0.03	41.10
Female 5	0.29 ± 0.09	31.25	3.59 + 1.01	28.14	0.66 ± 0.31	46.90	1.93 ± 1.77	91.76	0.28 ± 0.12	42.40	0.04 ± 0.02	48.78
Female 6	0.30 ± 0.11	36.67	3.83 ± 1.29	33.66	0.88 ± 0.67	76.14	2.21 ± 1.40	63.32	0.45 ± 0.23	50.77	0.09 ± 0.03	35.29

Table IV-2 Female urine metabolite variation over thirty days.

a. Six randomly selected urine metabolites of six female subjects over thirty days.

b. Normalized metabolite concentrations (([metabolite mM]/[Creatinine mM])*10), ± SD

c. Coefficient of variation 100*(SD/mean)

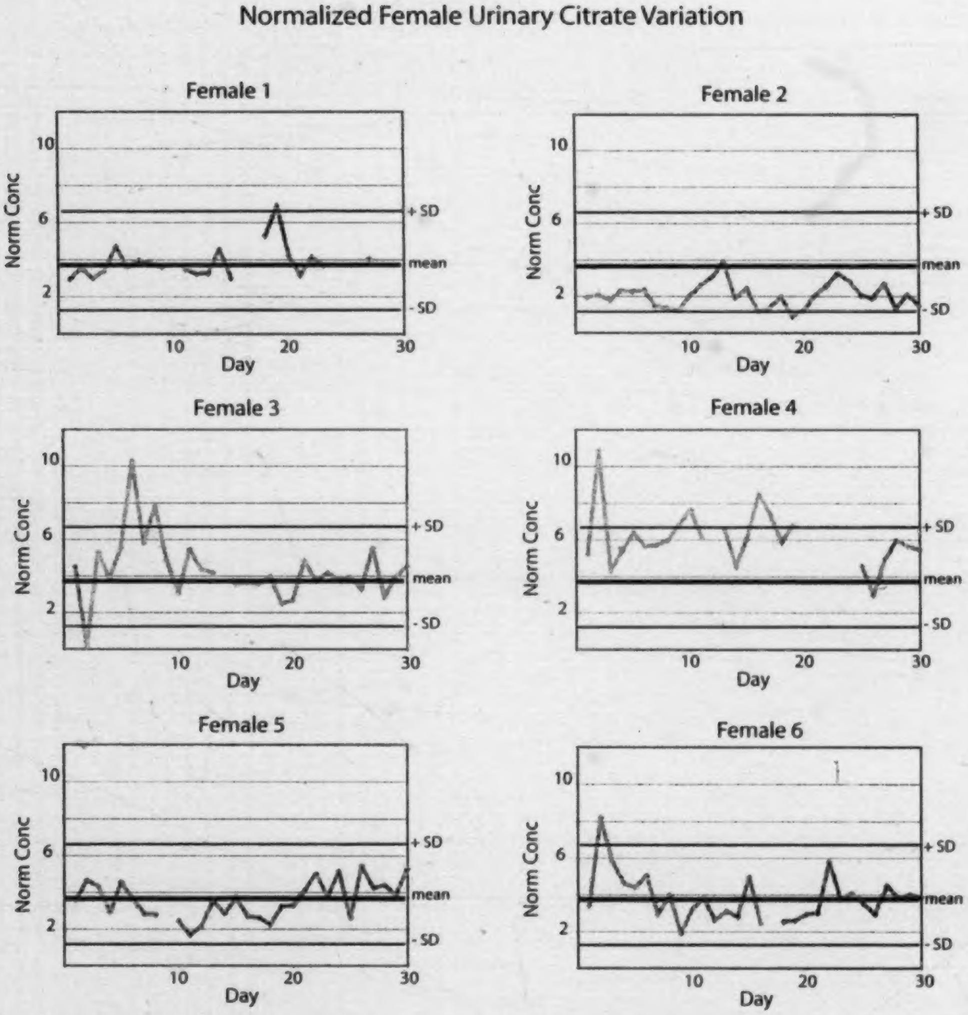


Figure IV-5. **Intra-individual urine citrate variability**. Normalized citrate levels for morning urine samples taken from six randomly selected female subjects over a period of 30 days. Thick black line denotes the female normal population mean (3.88); thinner black lines represent the population standard deviation (+/- 2.4).

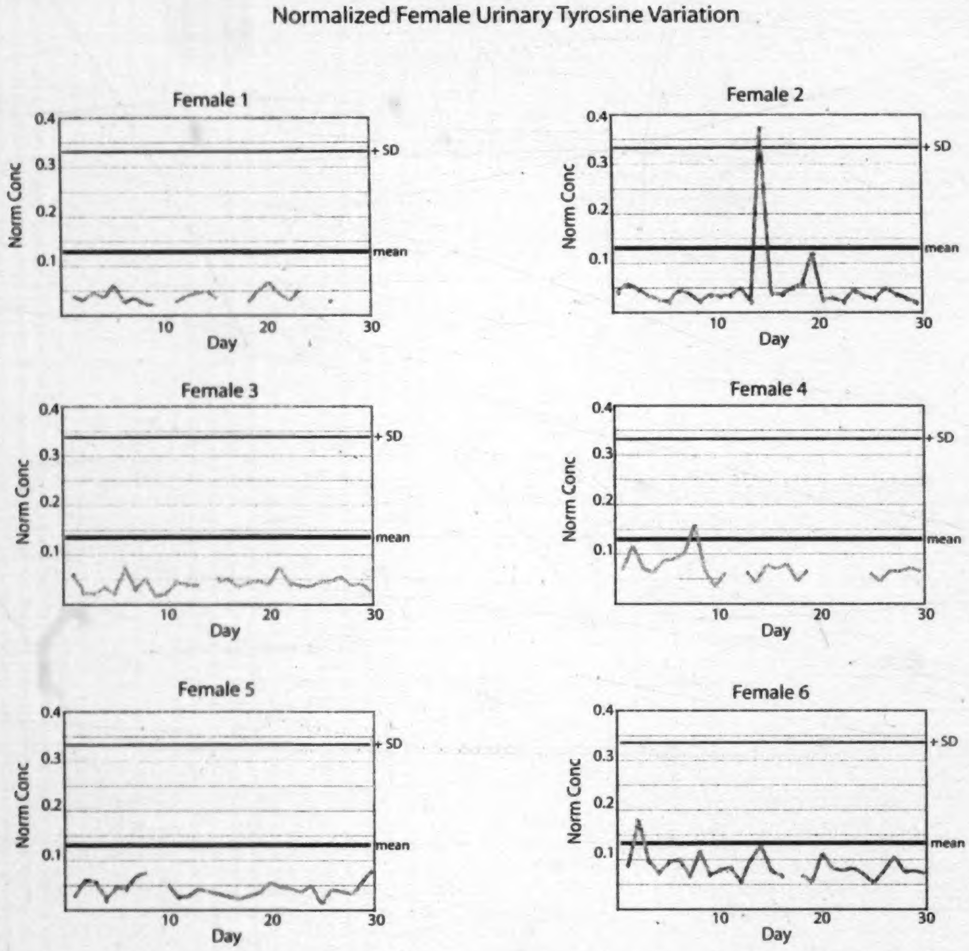


Figure IV-6. **Intra-individual urine tyrosine variability**. Normalized tyrosine levels for morning urine samples from six randomly selected female subjects over a period of 30 days. Thick black line denotes the female normal population mean (0.13); thinner black lines represent the population standard deviation (+/- 0.2).

V-H. Guinea Pig Variability

To investigate the possible influences of extraneous factors on the human population urine metabolite concentrations and variability (*e.g.* genetics, diet, stress, environment, etc.) 25 guinea pigs that were raised in a controlled environment were fed identical food, and were from similar genetic background, provided urine samples. Similar to the human data, the guinea pig urine creatinine levels varied considerably (2.85± 2.1mM) (Figure IV-7). The metabolites hippurate and citrate, chosen randomly before for the human normal population, were also quantified in guinea pig urine and are shown in Figures IV-8 and -9. Absolute urine hippurate concentrations for the guinea pig normal population are shown graphically in Figure IV-8 (a), (mean: 9.69 ± 9.0 mM). The normalized hippurate concentrations following individual normalization to creatinine (Eq.2) are shown in Figure IV-8 (b), (mean: 10.16 ± 7.3). For citrate the absolute concentration mean was 0.23 ± 0.42 (Figure IV-9a), and the normalized mean concentration was 0.57 ± 1.4 (Figure IV-9b). Analysis for 24 metabolites revealed similarly large standard deviations and coefficients of variation for both absolute and normalized concentrations for the guinea pig group (Table IV-3).

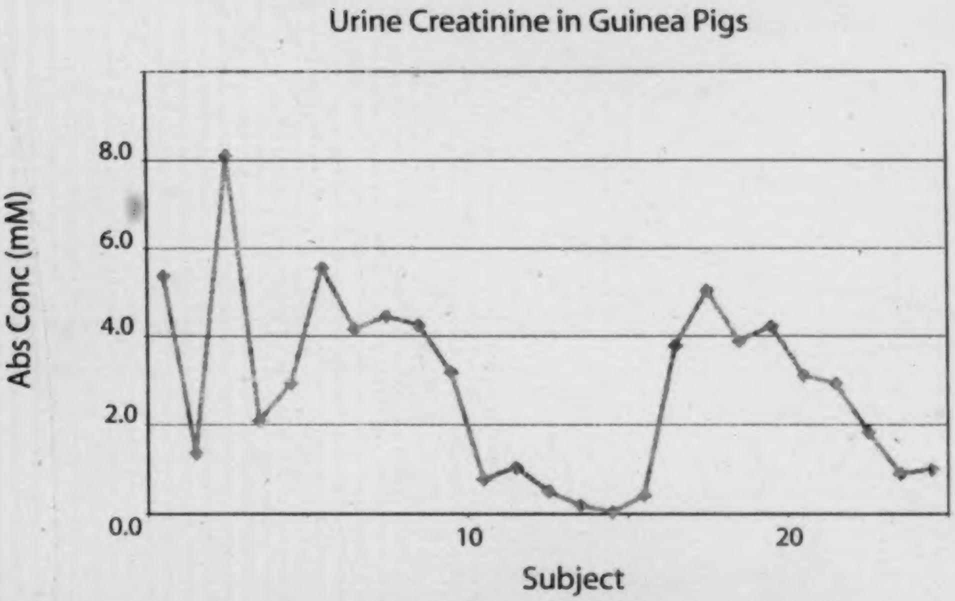


Figure IV-7. **Guinea pig urine creatinine variability**. Absolute creatinine concentrations of 25 control guinea pig urine samples (2.84 ± 2.05 mM).

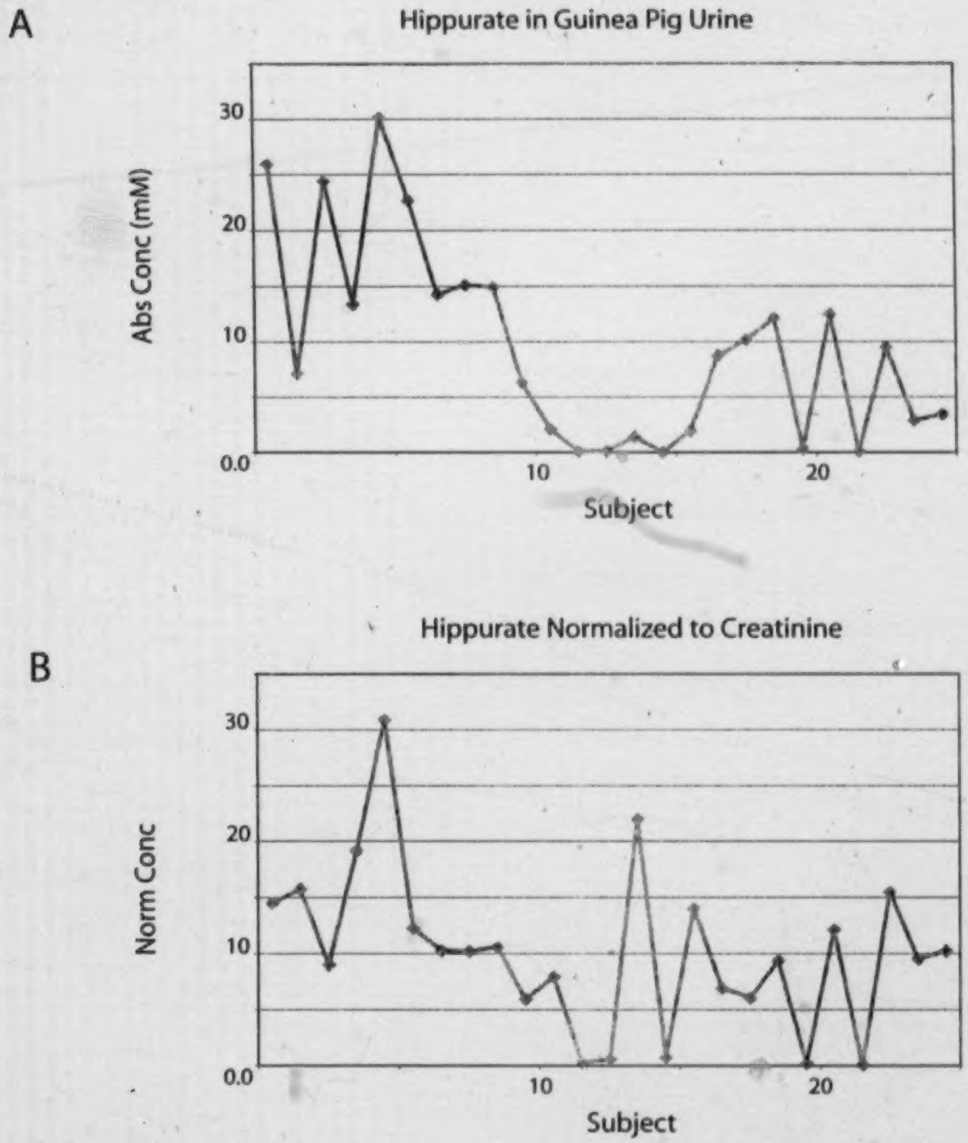


Figure IV-8. **Urine hippurate normalization in control guinea pigs.** Guinea pig urine hippurate concentrations prior to (9.58 ± 8.95 mM, A), and following normalization to creatinine (see Eq. 2., 10.2 ± 7.31,B).

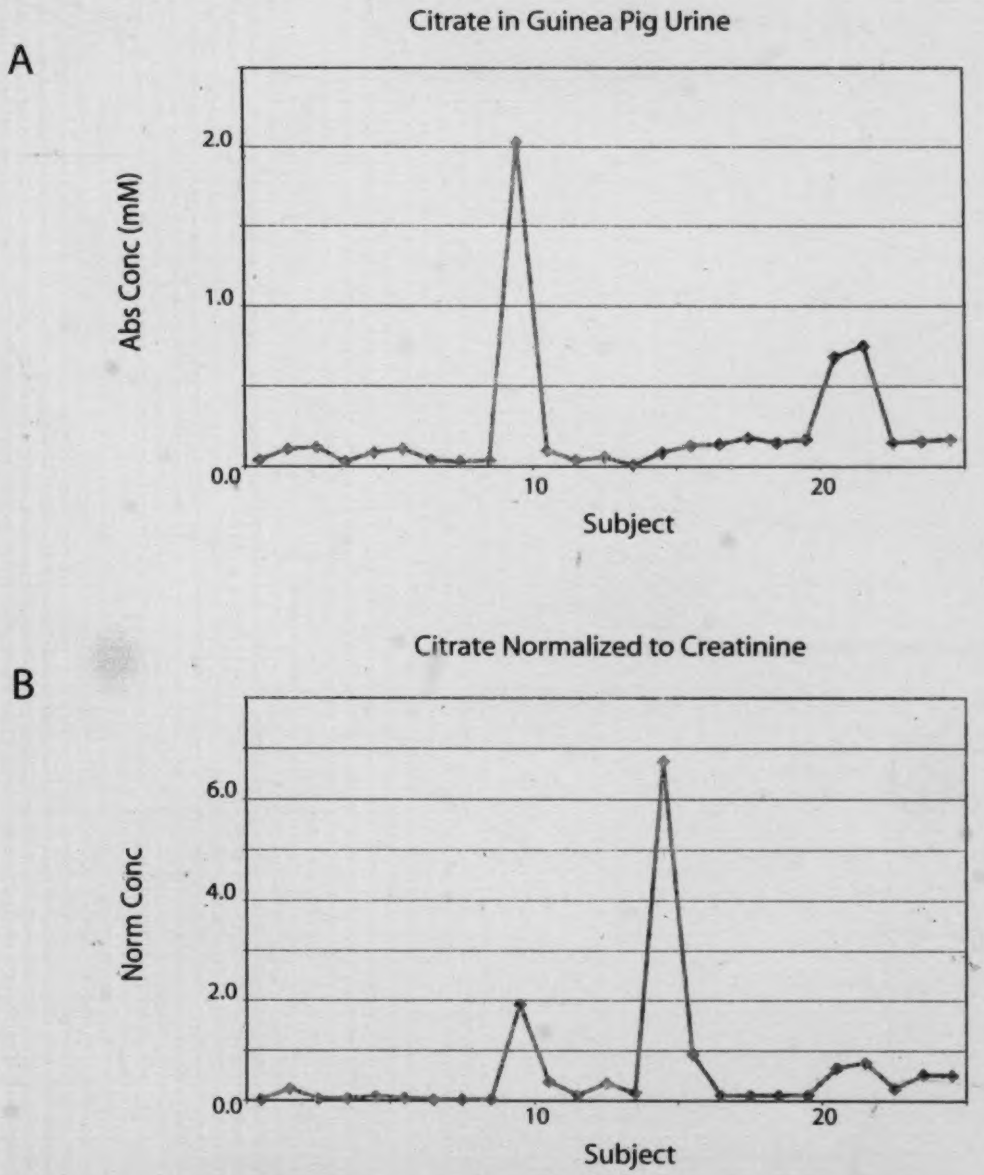


Figure IV-9. **Normalization of urine citrate in control guinea pigs**. Citrate levels in control guinea pig urine prior to (0.22 ± 0.42 mM, A), and following normalization to creatinine (see Eq. 2., 0.57 ± 1.35, B).

Metabolite	Absolute (mM)		Coeff Var	Normalized		Coeff Var
	mean ± SD	min/max		mean ± SD	min/max	
2-Hydroxyisobutyrate	0.13 ± 0.11	(0.01 - 0.37)	80.0	0.19 ± 0.14	(0.00 - 0.75)	76.2
2-Oxoisocaproate	0.05 ± 0.05	(0.00 - 0.18)	113	0.08 ± 0.15	(0.00 - 0.75)	175
4-Aminohippurate	0.03 ± 0.02	(0.00 - 0.06)	63.3	0.07 ± 0.15	(0.01 - 0.75)	212
4-Hydroxyphenylacetate	0.20 ± 0.25	(0.00 - 0.74)	122	0.23 ± 0.22	(0.00 - 0.75)	93.5
Alanine	0.16 ± 0.15	(0.06 - 0.69)	89.2	0.68 ± 2.21	(0.03 - 11.3)	323
Aspartate	0.08 ± 0.06	(0.00 - 0.20)	68.8	0.13 ± 0.14	(0.04 - 0.75)	105
Citrate	0.22 ± 0.42	(0.01 - 2.03)	185	0.57 ± 1.35	(0.02 - 6.75)	236
Creatine	0.13 ± 0.12	(0.02 - 0.62)	98.6	0.28 ± 0.58	(0.04 - 3.00)	207
Creatinine	2.84 ± 2.05	(0.04 - 8.11)	72.2	3.00 ± 0.00	(3.00 - 3.00)	
Cytidine	0.02 ± 0.03	(0.00 - 0.10)	132	0.06 ± 0.15	(0.00 - 0.75)	262
Formate	0.48 ± 0.43	(0.00 - 1.71)	88.7	2.37 ± 6.24	(0.00 - 30.8)	263
Glutamate	0.11 ± 0.08	(0.00 - 0.25)	68.8	0.21 ± 0.29	(0.00 - 1.5)	139
Hippurate	9.58 ± 8.95	(0.01 - 30.2)	93.4	10.2 ± 7.31	(0.06 - 30.9)	71.9
Histidine	0.00 ± 0.01	(0.00 - 0.03)	322	0.00 ± 0.00	(0.00 - 0.07)	327
Hypoxanthine	0.00 ± 0.01	(0.00 - 0.02)	136	0.01 ± 0.01	(0.00 - 0.06)	190
Lactate	0.27 ± 0.43	(0.03 - 2.06)	158	6.62 ± 30.8	(0.05 - 155)	465
N-Methylhistidine	0.00 ± 0.00	(0.00 - 0.00)	0.00	0.00 ± 0.00	(0.00 - 0.00)	0
Phenylalanine	0.13 ± 0.11	(0.00 - 0.36)	83.4	0.18 ± 0.14	(0.00 - 0.75)	79.9
Salicylurate	0.06 ± 0.07	(0.00 - 0.27)	113	0.06 ± 0.05	(0.00 - 0.21)	94
Trans-Aconitate	0.02 ± 0.02	(0.00 - 0.06)	108	0.03 ± 0.04	(0.00 - 0.18)	135
Threonine	0.06 ± 0.05	(0.00 - 0.16)	81.9	0.20 ± 0.59	(0.00 - 3.00)	299
Tryptophan	0.06 ± 0.05	(0.00 - 0.16)	80.2	0.10 ± 0.14	(0.00 - 0.75)	134
Tyrosine	0.33 ± 0.31	(0.02 - 1.07)	94.8	0.41 ± 0.44	(0.07 - 2.25)	107
Uridine	0.02 ± 0.02	(0.00 - 0.07)	111	0.05 ± 0.15	(0.00 - 0.75)	273

Table IV-3. Guinea pig urine metabolite concentrations

a. Metabolites were quantified for variability in normal guinea pigs (mM)

b. Creatinine was used for urine normalization (Eq. [2])

c. Buffer constituents DSS and imidazole were 0.49 and 10.0 mM respectively

d. Coefficient of variation 100*(SD/Mean)

DISCUSSION and CONCLUSION

The results from this study have important implications for those interested in examining urine for future metabolomic research. First, prior to a study describing a diseased state, as well as the unique properties and metabolic differences associated with the pathophysiology, it is important to define a 'normal, or control population'. Often studies compare different test or disease subjects to a normal cohort of a similar sample size. The difficulty is that the limited sample size, particularly when dealing with human clinical studies, can influence the conclusions of the study. Moreover, without examining trends and variation over time, single comparison of control values can generate misleading results.

In this study, questions were raised regarding the sampled population, and the philosophical definition of a "healthy, or normal human population". My inclusion criteria rested upon a physician's judgment that every subject had no underlying chronic illness or major medical problem, and a general understanding that the subject was 'healthy'. We intentionally avoided choosing 'idealized' normal subjects, as this would not represent an actual 'normal' population (*i.e.* most people use some pharmaceuticals, and very few are varsity athletes). No attempt was made to normalize diet, or other factors in order to best determine the actual variation present in normal individuals. Any additional attempt to regulate extraneous factors like sleep or exercise would falsely skew results towards an unattainable clinical scenario. Most previously published metabolomic and clinical studies that compare disease and control cohorts have not insisted upon a regimented 'control' group with regulated sleep, diet, exercise, and stress. Therefore, my study sampled a more inclusive normal population that better reflected the majority of the human control subjects

^{1}H-NMR spectroscopy was used to identify and quantitate metabolites in urine samples from a normal human cohort to determine the metabolite variability observed at the level of the population and the individual. A key finding of this study was that the variance (standard deviation and coefficient of variation) of the pooled metabolite concentrations within the normal population was high, as was the degree of variation of the metabolites for individuals over time. It is clear that individual metabolite profiles, as sampled in the urine, are capable of impressive variation and a clearer understanding of metabolite variability within control human subjects is needed in future metabolomic and clinical investigations.

Variation in urine volume as a result of dietary liquid intake has a direct effect on absolute metabolite concentrations. Previously, volume effects have been normalized by expressing metabolite concentrations as ratios relative to the endogenous creatinine levels for each urine sample. Creatinine is associated with overall muscle mass and body size and current literature states that creatinine exhibits little concentration variability related to normal biological function[31]. Clinically, serum creatinine is often used as an indicator of kidney function. Therefore, it is reasonable to assume creatinine offers the best possible internal standard for correcting urine volume effects. Table IV-1 and Figures IV-3 and -4 demonstrate that the standard deviations and coefficients of variation for absolute and normalized (Eq. [1]) metabolite concentrations in the human control group are similar. Therefore, volume effect (subject hydration) is not the major contributing factor to the observed variance in metabolite concentrations. The observed large coefficients of variation (up to 270 for the metabolite lactate) for normalized metabolite concentrations within the study's normal population was of interest because studies often attempt to diagnosis pathological metabolic profiles based upon limited population information (*e.g.* small sample size). The data shows that large variance in urinary metabolite concentrations in the normal human group is not due to urine volume effects, but rather reflects actual metabolite variance. Adequate knowledge and understanding of the variance of urinary metabolite concentrations in the normal human population is necessary for the interpretation of future metabolomic studies.

It is possible that the large population variance may be due to inter-subject variance (differences in individual mean values) due to gender, age, or diet, while each individual maintains a tight biological homeostasis[32]. From six random healthy female volunteers 10 metabolites were monitored over a 4 week period. The data demonstrated that individual citrate and tyrosine levels are capable of ranging through the full standard deviation of the entire population. Similar degrees of variability were observed for the other metabolites studied (Table IV-2, and similarly for the normal cohort, Table IV-1). This study clearly shows that a single individual is capable of varying greatly, and is able to span the same metabolite range as that for an entire group.

When we followed individual subjects over the 30 day period some metabolite 'spikes' were observed in some subjects (Figures IV-5 and 6). No correlation was found with the changes in metabolite concentrations and changes in the subjects' daily routine. However, female #1 had an increase in citrate excretion immediately before the beginning of her menstrual cycle. Although this only occurred for the one subject, it may point to an area requiring further research[33].

This study also chose to look at sampling time because of the infrequency of collection observed in actual clinical settings. Although morning sampling is typically sought there are many instances where this is not possible. Therefore, it was important to determine whether there was any urinary metabolite variation that resulted from differences in the timing of sample collection. Of the 24 metabolites examined during this study, the concentrations of nine were statistically different between morning and afternoon urine collection. Thus, differences in metabolite excretion should be noted during metabolic studies using urine as the biofluid. Urine metabolite variability has been seen in earlier metabolomic studies; such as public health and exposure limits to certain agents[34]. The metabolite variability observed in this human study is similar to that found in studies by Tate[12] and Holmes[35].

Twenty-five control guinea pigs were sampled to identify possible sources for the large metabolite variability identified in the normal human population. The guinea pigs were of similar genetic background, were raised in in a controlled environment, and were fed identical diets. Our data demonstrated that the guinea pig urine metabolites varied significantly (Figures IV-7-9 and Table IV-3). Interestingly the coefficients of variability for the metabolites identified were similar, and in some cases greater, when compared to the human control population. Many studies choose to use animal models because many of the extraneous factors identified as possible sources of error in human studies can be controlled. This study demonstrates that similar variation is possible in a control guinea pig population. For example, control rat urine metabolites have varied during estrus cycles[33], animal species [6, 36], diet [37], and even metabolic variation within similar species [12, 35]. A recent article in Science identifies the remarkable diversity in genetically identical cells, even following identical histories of environmental exposures[32].

Data from this study supported the conclusions of other human urine studies, which found that there was a significant difference in metabolite excretion between male and female subjects. My data shows that half of the metabolites examined demonstrated a statistically significant difference between sex groups. It is important that investigators choose the appropriate gender to compare between test and control populations (*i.e.* both populations of same sex, or equal mixture of gender) since I have shown that differences in metabolite excretion, as a result of gender, may influence metabolic conclusions.

Although studies for urine metabolites in human control subjects are limited, there are studies which attempt to find correlation of urinary metabolites with age and diet [38], variation within large populations [34, 39], and variation of metabolites over time [24, 40]. Investigators have identified high inter-individual variations in urine metabolites, while intra-individual variation was low

[24]. These investigators concluded that most subjects had a characteristic profile that did not change with time and was independent of diet. Those results are not consistent with the degree of metabolite variation observed in this study. Although the average metabolite concentrations and standard deviations were strikingly similar for the normal population compared with Zuppi et al. [24], the large degree of variability was not reduced when single individuals were followed over time in this study. Since Zuppi et al. collected urine only once a month for four months; the difference in collection frequency may explain the discrepancy. In addition, the larger population sampled in this study did not reduce the degree of variability. This suggests a unique population distribution that should be considered when performing the most basic statistical analysis. Other investigators have found that although there was high inter-individual variability, the intra- individual variability remained small [40]. Of the 24 metabolites measured in this study, I found high inter- *and* intra-individual variability, as well, the different findings may be the result of the limited diet and short duration of the Lenz study. Unfortunately, they do not report actual metabolite concentrations making comparison to this study difficult.

Despite the intra-individual variability between our study and Zuppi et al., the population averages for three of the four metabolites (*e.g.* hippurate and citrate) reported in their study were strikingly similar to the population averages found in my study. Given the geographic difference (Rome, Italy for Zuppi [24]vs. Edmonton, Canada for this study), and the probable genetic and diet differences between subjects in our study and those from Italy, the similarity in population averages between four metabolites from the two studies is intriguing.

Urine is not a fluid in constant circulation throughout the body that requires monitoring like the blood. Rather, urine is a collection of waste and biological by-products that are reflective of a larger mixture of metabolic processes that may have occurred over a long time period. The large variability in metabolite concentrations within the control subjects may be a reflection of the overall biological function of urine. However, the ease of urine sample collection has lead to its widespread use as the biofluid for metabolomic studies of many different human disease states. This study has highlighted important factors such as metabolite excretion differences due to sampling time, gender, and inter- and intra- individual variability; each of which impact the growing use of urine in metabolomic and future clinical diagnostic studies. This investigation demonstrates the great care that must be taken when establishing normal metabolite baselines for comparison in basic research or clinical investigations.

LITERATURE CITED

1. Kraut, J.A. and I. Kurtz, Metabolic acidosis of CKD: diagnosis, clinical characteristics, and treatment. Am J Kidney Dis, 2005. 45(6); 978-93.
2. Trachtenbarg, D.E., Diabetic ketoacidosis. Am Fam Physician, 2005. 71(9); 1705-14.
3. Shockcor, J.P. and E. Holmes, Metabonomic applications in toxicity screening and disease diagnosis. Curr Top Med Chem, 2002. 2(1); 35-51.
4. Kubitz, R., V. Keitel, and D. Haussinger, Inborn errors of biliary canalicular transport systems. Methods Enzymol, 2005. 400; 558-69.
5. Green, R.H., et al., Asthma exacerbations and sputum eosinophil counts: a randomised controlled trial. Lancet, 2002. 360(9347); 1715-21.
6. Bollard, M.E., et al., NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. NMR Biomed, 2005. 18(3); 143-62.
7. Moolenaar, S.H., U.F. Engelke, and R.A. Wevers, Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism. Ann Clin Biochem, 2003. 40(Pt 1); 16-24.
8. Weckwerth, W., Metabolomics in systems biology. Annu Rev Plant Biol, 2003. 54; 669-89.
9. Rastall, R.A., et al., Modulation of the microbial ecology of the human colon by probiotics, prebiotics and synbiotics to enhance human health: an overview of enabling science and potential applications. FEMS Microbiol Ecol, 2005. 52(2); 145-52.
10. Keller, N.P., G. Turner, and J.W. Bennett, Fungal secondary metabolism - from biochemistry to genomics. Nat Rev Microbiol, 2005. 3(12); 937-47.
11. Singer, M., Metabolic failure. Crit Care Med, 2005. 33(12 Suppl); S539-42.
12. Tate, A.R., S.J. Damment, and J.C. Lindon, Investigation of the metabolite variation in control rat urine using (1)H NMR spectroscopy. Anal Biochem, 2001. 291(1); 17-26.

13. Coen, M., et al., Proton nuclear magnetic resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis. Clin Infect Dis, 2005. 41(11); 1582-90.

14. Duarte, I.F., et al., Metabolic assessment of human liver transplants from biopsy samples at the donor and recipient stages using high-resolution magic angle spinning 1H NMR spectroscopy. Anal Chem, 2005. 77(17); 5570-8.

15. Stanley, E.G., et al., Sexual dimorphism in urinary metabolite profiles of Han Wistar rats revealed by nuclear-magnetic-resonance-based metabonomics. Anal Biochem, 2005. 343(2); 195-202.

16. Sabatine, M.S., et al., Metabolomic identification of novel biomarkers of myocardial ischemia. Circulation, 2005. 112(25); 3868-75.

17. Wennergren, G., Inflammatory mediators in blood and urine. Paediatr Respir Rev, 2000. 1(3); 259-65.

18. Brindle, J.T., et al., Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med, 2002. 8(12); 1439-44.

19. van Rhijn, B.W., H.G. van der Poel, and T.H. van der Kwast, Urine markers for bladder cancer surveillance: a systematic review. Eur Urol, 2005. 47(6); 736-48.

20. de Jongste, J.C., Surrogate markers of airway inflammation: inflammometry in paediatric respiratory medicine. Paediatr Respir Rev, 2000. 1(4); 354-60.

21. Stoyanova, R. and T.R. Brown, NMR spectral quantitation by principal component analysis. NMR Biomed, 2001. 14(4); 271-7.

22. Sardari, S. and D. Sardari, Applications of artificial neural network in AIDS research and therapy. Curr Pharm Des, 2002. 8(8); 659-70.

23. Gavaghan, C.L., I.D. Wilson, and J.K. Nicholson, Physiological variation in metabolic phenotyping and functional genomic studies: use of orthogonal signal correction and PLS-DA. FEBS Lett, 2002. 530(1-3); 191-6.

24. Zuppi, C., et al., 1H NMR spectra of normal urines: reference ranges of the major metabolites. Clin Chim Acta, 1997. 265(1); 85-97.

25. Keun, H.C., et al., Analytical reproducibility in (1)H NMR-based metabonomic urinalysis. Chem Res Toxicol, 2002. 15(11); 1380-6.

26. Geigy, a.L., Geigy scientific tables. 8th, rev. and enl. ed. ed, ed. C. Lentner. Vol. 1. 1981 - 1992, West Cadwell, NJ: Ciba-Geigy Corp.

27. Chung, Y.L., et al., Muscle metabolites, detected in urine by proton spectroscopy, correlate with disease damage in juvenile idiopathic inflammatory myopathies. Arthritis Rheum, 2005. 53(4); 565-70.

28. Lenz, E.M., et al., Metabonomics, dietary influences and cultural differences: a 1H NMR-based study of urine samples obtained from healthy British and Swedish subjects. J Pharm Biomed Anal, 2004. 36(4); 841-9.

29. Kumar, A., R.R. Ernst, and K. Wuthrich, A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem Biophys Res Commun, 1980. 95(1); 1-6.

30. Saude, E.J., C.M. Slupsky, and B.D. Sykes, Optimization of NMR analysis of biological fluids for quantitative accuracy. Metabolomics, 2006. 2(3); 113-23.

31. Wyss, M. and R. Kaddurah-Daouk, Creatine and creatinine metabolism. Physiol Rev, 2000. 80(3); 1107-213.

32. Raser, J.M. and E.K. O'Shea, Noise in gene expression: origins, consequences, and control. Science, 2005. 309(5743); 2010-3.

33. Bollard, M.E., et al., Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution (1)H NMR spectroscopy of urine and pattern recognition. Anal Biochem, 2001. 295(2); 194-202.

34. Symanski, E. and N.M. Greeson, Assessment of variability in biomonitoring data using a large database of biological measures of exposure. AIHA J (Fairfax, Va), 2002. 63(4); 390-401.

35. Holmes, E., et al., Chemometric models for toxicity classification based on NMR spectra of biofluids. Chem Res Toxicol, 2000. 13(6); 471-8.

36. Ebbels, T.M., et al., Evaluation of metabolic variation in normal rat strains from a statistical analysis of 1H NMR spectra of urine. J Pharm Biomed Anal, 2004. 36(4); 823-33.

37. Phipps, A.N., et al., Effect of diet on the urinary excretion of hippuric acid and other dietary-derived aromatics in rat. A complex interaction between diet, gut microflora and substrate specificity. Xenobiotica, 1998. 28(5); 527-37.

38. Guneral, F. and C. Bachmann, Age-related reference values for urinary organic acids in a healthy Turkish pediatric population. Clin Chem, 1994. 40(6); 862-6.

39. Dyer, A., et al., Urinary biochemical markers of dietary intake in the INTERSALT study. Am J Clin Nutr, 1997. 65(4 Suppl); 1246S-53S.

40. Lenz, E.M., et al., A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. J Pharm Biomed Anal, 2003. 33(5); 1103-15.

CHAPTER V

Metabolomic investigation of cellular biochemistry and pulmonary pathophysiology

OVERVIEW

Disorders in the respiratory system, such as cystic fibrosis (CF), involve the infiltration and activation of airway inflammatory cells, including neutrophils. The infiltration and activation of the neutrophil leads to the secretion of peroxidases, which react further with substrates in solution to produce oxidative metabolites such as 3-chlorotyrosine. Elevation of modified tyrosine residues in the airways of patients with CF may be detectable by NMR spectroscopy in correlation with inflammatory cell influx. *This chapter demonstrates the possibility of correlating measured metabolites in human sputum to a unique clinical pathophysiological state through metabolomic analysis by NMR.* High-resolution ^{1}H-NMR spectroscopy was used to analyze the production of modified tyrosine residues from *in vitro* stimulation of peripheral blood eosinophils and neutrophils, as well as in sputum samples from control subjects and patients with CF. The *in vitro* stimulation of the purified peripheral blood neutrophils generated 3-chlorotyrosine, while eosinophils produced predominantly 3-bromotyrosine and 3,5-dibromotyrosine. Similar metabolites were identified and quantified in sputum samples taken from CF patients. In addition, cell counts from the sputum samples were correlated with the generation of modified tyrosine residues. This study was important for demonstrating the applicability of NMR analysis for clinical metabolomic investigations.

INTRODUCTION

Cystic fibrosis is an autosomal recessive disorder caused by mutations in the CF transmembrane regulator (CFTR) gene and is the most common serious genetic disease in Caucasian populations[1]. Evidence suggests that the lungs are histologically normal at birth and pulmonary damage is initiated later due to altered epithelial Cl^- and water transport, increased mucus viscosity, reduced mucociliary clearance, and decreased antibacterial defense within the respiratory tract[2]. Patients with the disease become increasingly susceptible to bacteria-induced respiratory tract infections, and often die of respiratory failure due to repeated acute pulmonary infections[1-3]. Bronchial secretions from CF patients are commonly enriched in neutrophils and contain high amounts of neutrophil-derived mediators including the neutrophil granule enzyme myeloperoxidase (MPO) [3-5].

Stimulation of neutrophils leads to the activation of superoxide-generating NADPH oxidase as well as the release of MPO by degranulation[6-8]. Superoxide produced from NADPH oxidase is rapidly dismutated to form H_2O_2, which is utilized by MPO along with physiological concentrations of Cl^- to produce hypochlorous acid (HOCl). HOCl, the active ingredient of bleach, is a crucial mediator in microbial killing, and can also evoke oxidative damage to the various cellular or extracellular constituents within inflamed tissues, including oxidative modification of free tyrosine leading to the formation of 3-chlorotyrosine and its derivatives[9].

A key granule protein and an important effector molecule released during eosinophil degranulation is the enzyme eosinophil peroxidase (EPO), which is stored in the matrix of the eosinophil crystalloid granule[10-12]. Eosinophils also undergo respiratory burst due to the activation of NADPH oxidase. EPO preferentially interacts with Br^- along with H_2O_2 to generate hypobromous acid (HOBr), even in physiological conditions where Cl^- and other halide ions are present in much greater concentrations[10]. Wu et al [13], and others [14] have demonstrated the propensity of hypobromous acid to target numerous compounds, including the amino acid tyrosine, leading to the production of 3-bromotyrosine and its derivative, 3,5-dibromotyrosine.

Earlier studies have utilized GC, MS, and HPLC to analyze the presence of modified tyrosine residues in biological specimens[9, 13, 14]. This study set out to investigate the feasibility of applying high-resolution NMR spectroscopy to measure modified tyrosine residues in peripheral blood eosinophils and neutrophils, and to determine whether brominated and/or chlorinated tyrosine residues can be detected in induced sputum samples from patients with CF. The capacity of NMR to detect a wide range of biologically relevant compounds from samples that require little to no preparation and are not chemically modified adds to the attractiveness of NMR as a tool for medically relevant investigations.

The use of induced sputum, a relatively non-invasive technique, has shown excellent reproducibility [15, 16] and validity in its measurement of inflammatory changes in the lower airways [15-18]. This study demonstrates that levels of 3-chlorotyrosine are elevated relative to control subjects in sputum samples from CF patients, and that these increased levels correlate with the numbers of sputum neutrophils. While 3-bromotyrosine was nearly always undetectable in control sputum samples, it was present in CF sputum, suggesting that eosinophil activation may occur in CF. These findings confirm that high resolution NMR spectroscopy is a valid tool for the analytical investigation of *in vitro* biological experiments and *in vivo* samples taken directly from patients.

EXPERIMENTAL PROCEDURES

V-A. Volunteers

Informed consent was obtained from non-smoking volunteers in accordance with guidelines established by the University of Alberta Health Research Ethics Board.

V-B. Isolation and Purification of Eosinophils and Neutrophils

To isolate neutrophils, 50-100 ml of peripheral blood was obtained from normal (non-atopic) subjects and subjected to erythrocyte sedimentation, followed by density centrifugation on Ficoll, resulting in high neutrophil purity (average

>98%). Eosinophils were purified from the peripheral blood as previously described[19]. Briefly, peripheral blood samples (100 ml) were obtained from mild atopic asthmatic and atopic non-asthmatic subjects displaying eosinophilia >2%, who were not receiving oral corticosteroids[19, 20]. Whole blood was subjected to erythrocyte sedimentation in 6% dextran and upper phase cells were centrifuged on a single-step Ficoll gradient before highly purified $CD16^-$ eosinophils (>99%) were isolated using negative immunomagnetic selection.

V-C. NMR Spectral Analysis of Resting and Stimulated Cells

To test the ability of high-resolution NMR spectroscopy to identify indices of cellular stimulation *in vitro* isolated neutrophils (4×10^6) were placed in a 1.5-ml Eppendorff tube containing 550 μl of NMR buffer solution (Dulbecco's PBS, pH 7.4, 5% D_2O (Sigma-Aldrich, Mississauga, ON), 1 mM disodium-2, 2-dimethyl (2-silapentane-5-sulphonate) (DSS, Sigma-Aldrich), and 10 mM imidazole (Isotech, Matheson, OH)) supplemented with 1 mM L-tyrosine (Sigma-Aldrich), 100 μM NaBr, and 1.75 mM $CaCl_2$. The sample was transferred to a standard 5 mm-glass NMR sample tube. Cellular sedimentation and oxygenation was not a factor given the short spectral acquisition time (less than 10 minutes). All NMR spectra were acquired on a 500 MHz Varian Inova spectrometer equipped with a 5 mm triple resonance probe with actively shielded Z-gradient (NANUC, University of Alberta, Edmonton). One-dimensional ^1H spectra were collected at 25°C with a spectrum width of 8000 Hz, an acquisition time per scan of 4 s, 16 steady state scans preceding acquisition, and 128 transients were acquired for each spectrum, which was apodized with an exponential decay corresponding to a line broadening of 0.5 Hz prior to Fourier Transformation. Buffer constituents DSS and imidazole served as internal standards for spectral quantification and sample pH determination, respectively. The methyl resonance peak of DSS was utilized as an internal reference set to 0 ppm. Spectral analysis and chemical identification was determined according to known chemical shifts, scalar couplings, and by sample spiking with pure standards. Spectral quantification of chemical constituents utilized a VNMR 6.1C software package (Varian Inc, Palo Alto, CA). Areas for spectral integration were carefully chosen to include resonant peaks of interest while excluding surrounding signals. Integration regions were maintained for similar resonant peaks between patients. Following an initial scan of control unstimulated neutrophils a stimulatory process was applied, involving a pre-incubation with cytochalasin B (5 μg/ml) at 37°C for 5 minutes followed by the addition of 5 μM f-Met-Leu-Phe (fMLP) for at least 15 min, followed by a second NMR scan (nt=1024). Isolated eosinophils (3 x 10^6) underwent a similar procedure, but instead were stimulated by the addition of a cytokine cocktail (10 ng/ml each of interleukin (IL)-3, IL-5, and granulocyte/macrophage colony-stimulating factor (GM-CSF)).

V-D. Sputum Collection and Processing

To test the ability of high resolution NMR spectroscopy to identify indices of cellular stimulation *in vivo*, patients were recruited through the University of Alberta Hospital asthma clinic (Table V-1). The CF subjects were chronically colonized by *Pseudomonas aeruginosa* but did not have an acute infection and were receiving neither systemic nor inhaled corticosteroids at the time of sputum collection. As controls, sputum samples were collected from nine life-long non-smoking subjects of similar age. For all subjects, pretreatment with 200 μg of salbutamol was administered to reduce the risk of bronchoconstrictive episodes during sputum induction. Baseline spirometry was obtained using American Thoracic Society standards and was measured during sputum induction to ensure safety of the procedure[15, 16]. The sputum induction procedure was done as previously described[15, 16]. Briefly, an ultrasonic nebulizer was used to deliver normal (0.9% NaCl) and hypertonic sterile saline solutions at 3%, 4%, or 5% NaCl. Each solution was inhaled for a duration of 7 minutes. The induction procedure was terminated when the subject's FEV_1 (forced expiratory volume in one second) fell ≥10% from baseline with normal saline or ≥ 20% with hypertonic saline. Subjects were required to cough deeply to produce a suitable sample for testing. After each interval of saline inhalation, subjects were instructed to clear their throat and nasal passages in order to prevent any contamination by upper airway or oral secretions. Plugs of sputum were manually separated from expectorate and stored free of preservatives at –80°C. Sample identification was blinded with an accession number including the date of collection to coordinate with the subject's induction and differential evaluation.

Sample	N	Age (avg.)	Sex	Eosinophil (% cell pop)	Neutrophil (%cell pop)	Pre-Bronchodilator (FEV1/FVC)
Control	9	28±5	4 male 5 female	1.3 ± 0.4	13.3 ± 6.8	83.0 ± 2.4
Cystic Fibrosis	7	29±13	4 male 3 female	0.64 ± 0.6	97.1 ± 2.0	53.8 ± 4.9

Table V-1. Patient characteristics, sputum cell counts, and spirometry

V-E. Cell Counts

Cytospin preparations were prepared from sputum samples using methods previously described[15, 16], and were counted by a technician blinded to the source of the samples.

V-F. NMR Spectroscopy of Sputum

Immediately before analysis, 90 mg of sputum was placed in a 1.5-ml Eppendorff tube containing 550 µl of NMR buffer solution supplemented with 2.5 mM NaN_3. The sample was subjected to sonication (single burst of 10 s at 12 W) using a micro-tip sonicator (Fisher Scientific, Mississauga, ON). The sample (600 µl) was removed from the Eppendorff tube and placed in a standard 5 mm-glass NMR sample tube (Wilmad-Labglass, Buena, NJ) for spectral analysis. Spectra were collected and analyzed in the same manner as described above, with the exception that 512 transients were acquired for each spectrum.

V-G. Statistical Analysis

A subset of both sputum cell counts and NMR-measured biomarker variables were chosen to reduce the dimensionality of analysis. Forward and backward stepwise regression analyses were used. This resulted in retaining both cell count variables in the induced sputum (neutrophil, eosinophil, macrophages, and lymphocytes), and NMR-detected variables (3-bromotyrosine, 3-chlorotyrosine, and 3,5-dibromotyrosine). NMR also identified dityrosine according to known chemical shifts, scalar couplings, and sample spiking with pure standards. Since dityrosine can be formed as a result of both eosinophil and neutrophil stimulation no significant correlations were found between dityrosine formation and either eosinophil or neutrophil stimulation *in vivo*.

The primary analysis consisted of finding significant associations between clinical and NMR variables. A correlation matrix containing all pair-wise correlations between clinical and NMR variables for the CF population was computed. For outcome variables of control and CF populations, a two-sided nonparametric Mann-Whitney U test was used. Data was considered statistically significant when the analysis generated $p < 0.05$.

RESULTS

V-A. Chlorination of Free Tyrosine Residues

To investigate whether neutrophils produce chlorotyrosine residues during in vitro stimulation in the presence of free tyrosine and if these oxidative metabolites are detectable by NMR, I examined chlorotyrosine formation by NMR analysis. Physiological concentrations of halides (Cl^- and Br^-) were used[10], and although physiological concentrations of tyrosine were visible by NMR tenfold higher concentrations were used to ensure sufficient free tyrosine for the chemical reactions[21]. Purified neutrophils were preincubated with cytochalasin B (5 μg/ml) for 5 min followed by fMLP (5 μM) in NMR buffer supplemented with physiological concentrations of Cl^-, Br^- and 1 mM L-tyrosine. It was found that neutrophils preferentially generated 3-chlorotyrosine following the release of MPO and oxidative metabolites (Figure IV-1a). The specificity of MPO for chloride ions was demonstrated by the identification of chlorinated tyrosyl-rings in the presence of physiologically relevant levels of other halide ions. Identification of 3-chlorotyrosine (7.086 and 7.305 ppm) was made by comparison of NMR chemical shifts and coupling constants with those from a known standard of 3-chlorotyrosine (Figure V-1b).

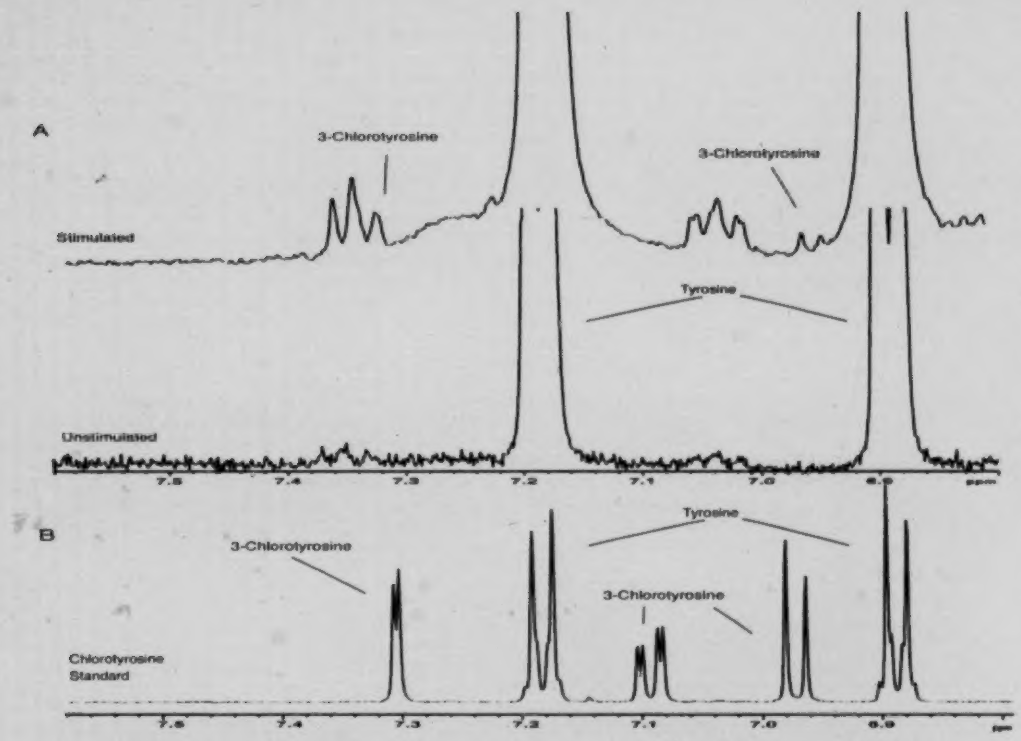


Figure V-1. **Formation of 3-chlorotyrosine by stimulated neutrophils.** *In vitro* formation of 3-chlorotyrosine in fMLP-stimulated neutrophils. One-dimensional ^1H NMR spectra (phenolic region) of isolated blood neutrophils ($4x10^6$, nt=128, lb=0.5 Hz). (A) Neutrophils were activated with cytochalasin B and fMLP, which led to the generation of 3-chlorotyrosine (nt=1024 lb=0.5 Hz). (B) For comparison, a spectrum of pure tyrosine and chlorinated tyrosine standards is also shown (nt=32, lb=0.5 Hz) (n=3).

Stimulation of neutrophils in a buffer lacking Cl^- resulted in no 3-chlorotyrosine production (data not shown). Slight frequency shifts were noticed upon comparison of pure chlorotyrosine standards with cell preparations. Such minor frequency shifts were attributed to pH differences between cell preparations during cellular activation and the pure standards. Chlorotyrosine identification was further confirmed by spiking samples with standard 3-chlorotyrosine. Later analysis also demonstrated the presence of 3,5-dichlorotyrosine (7.42ppm), but due to the lack of purified standards and difficulty in identification of this compound during sputum analysis, the study focused the analysis on 3-chlorotyrosine formation. Additional peaks are visible in Figures V-1 and -2, which arise from the one percent natural abundance ^{13}C-satellites of the tyrosine phenolic ring (7.35ppm, 7.02ppm and 7.06ppm, 6.72ppm). During integration the carbon satellites were omitted from the integration region, as they make only a minor contribution to overall quantitative accuracy. The amount of 3-chlorotyrosine produced by neutrophils during *in vitro* stimulation averaged 58 ± 7.1 mM 3-chlorotyrosine/10^6 neutrophils. These findings suggest that *in vitro* stimulation of neutrophils in the presence of physiological concentrations of Cl^- and Br^- results in the formation of only 3-chlorotyrosine, with undetectable levels of 3-bromotyrosine or 3,5-dibromotyrosine.

V-B. Bromination of Free Tyrosine Residues

This study determined the ability of NMR to measure the generation of brominated tyrosine residues by eosinophils *in vitro*. Samples of highly purified human eosinophils were buffered in a NMR solution supplemented with physiological concentrations of Cl^-, Br^- and 1 mM L-tyrosine. The samples were inserted into the core of a 500 MHz magnet and analyzed prior to stimulation. After the addition of a cytokine cocktail (10 ng/ml each of IL-3, IL-5, and GM-CSF), eosinophils were re-analyzed for spectral changes from the control baseline values. From this information I was able to readily identify the formation of free tyrosine residues modified at the 3' and 3', 5' positions of the tyrosyl phenolic ring by the addition of a bromide ion as early as eight minutes following stimulation of the isolated eosinophils with the cytokine cocktail (Figure V-2a). Identification of 3-bromotyrosine (resonance peaks at 6.965 and 7.130 ppm) and 3,5-dibromotyrosine (7.355 ppm) were made by comparison of the observed NMR chemical shifts and scalar-coupling constants to known standards, shown in Figure V-2b. NMR identification was further confirmed through the spiking of experimental samples with pure standards of brominated tyrosine residues. Stimulation of eosinophils in a buffer containing Cl^-, but no Br- resulted in no production of oxidatively modified free tyrosine residues (data not shown).

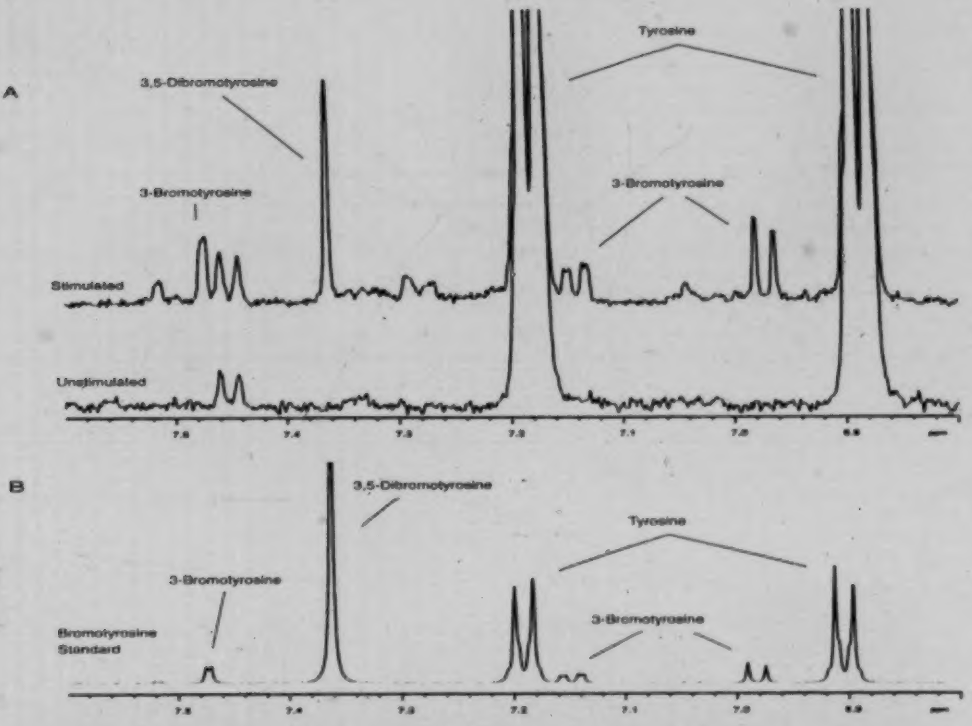


Figure V-2. **Formation of brominated tyrosine in stimulated eosinophils**. One-dimensional ^{1}H NMR spectra (phenolic region) of isolated blood eosinophils (3x10^6). The spectrum of unstimulated eosinophils demonstrates the presence of exogenous L-tyrosine, but lacks any evidence of oxidative modifications (nt=128, lb=0.5 Hz). The spectrum of stimulated (Il-3, IL-5, GM-CSF) eosinophils demonstrates cellular activation, and the subsequent production of (A) 3-bromotyrosine and 3,5-dibromotyrosine (nt=1024, lb=0.5 Hz). (B) The NMR spectrum of tyrosine and brominated tyrosine is also included (nt=32, lb=0.5 Hz).

An average of 1.5 mM ± 1.6 mM 3-bromotyrosine/10^6 eosinophils were generated in these experiments and are similar to previously published *in vitro* experiments [13] (4.5 mM ± 1mM 3-bromotyrosine/10^6 eosinophil). In a separate experiment, eosinophils were lysed with 0.1% Triton X-100 and incubated in the presence of bromide ions, L-tyrosine, and H_2O_2, which led to enzymatic production of HOBr, 3-bromotyrosine and 3,5-dibromotyrosine that could be detected by NMR analysis (data not shown). These findings suggest that eosinophils generate mainly 3-bromotyrosine and 3,5-dibromotyrosine, and not 3-chlorotyrosine, in the presence of physiological concentrations of Cl^- and Br^-, even though Cl^- is well in excess of Br^-.

V-C. Modified Tyrosine Residue Detection by NMR in Patient Sputum Samples

This study investigated the potential contribution of neutrophil- and eosinophil-specific peroxidase activity to the formation of modified tyrosine residues in sputum samples from control subjects in comparison with CF patients. As expected, cell counts were markedly different between the control subjects and CF patients, and agree with previously published results[22]. (Table V-1). Increased neutrophil counts were found in CF patients (97.1 ± 2.0%) when compared with control samples, while eosinophil counts were similar (Table V-1). NMR spectra of samples from the control subjects were compared with CF sputum samples (Figure V-3). This study observed many spectral differences between sputum samples from the two groups. Although the aliphatic regions of the spectra demonstrates visible difference between the CF and control groups this study deals with the modification of tyrosine residues and therefore the phenolic region where resonant peaks from tyrosine are easily identified. The phenolic region of the spectra was the focus of this analysis since these exhibited peaks are associated with free L-tyrosine and modified tyrosine residues (shown as insets in Figure V-3). Care was taken when interpreting the data due to the wealth of resonance peaks surrounding the biomarkers of interest. Peaks corresponding to free tyrosine were detected in both the control and CF sputum samples, while peaks corresponding to oxidatively modified tyrosine residues were found solely in CF sputum samples. Control subjects had no identifiable oxidatively modified tyrosine residues, which may correspond with their negligible eosinophil and neutrophil cell counts, and suggested a lack of inflammatory cellular activation.

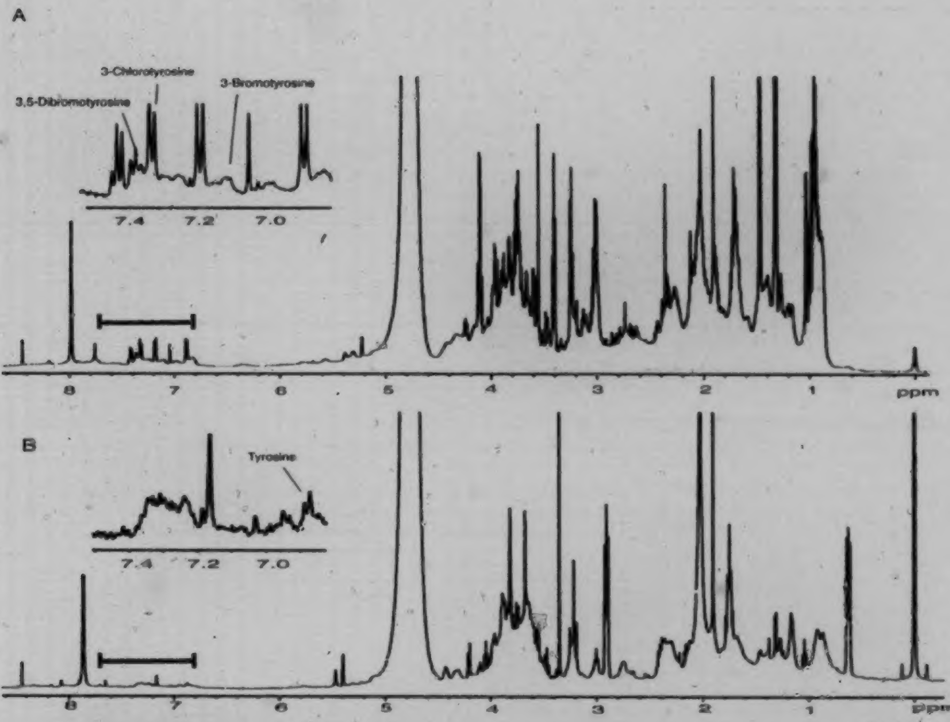


Figure V-3. **Spectra of sputum from CF and control subjects**. Full ^{1}H one-dimensional NMR spectra of (A) cystic fibrosis and (B) control sputum samples (nt=512, lb=0.5 Hz). Insets showing the phenolic regions (7.0-7.5 ppm) are also provided to demonstrate detectable levels of naturally occurring tyrosine and oxidized tyrosine produced as a result of eosinophil and neutrophil activation and peroxidase modification.

In agreement with previous findings, the oxidative marker dityrosine was observed in sputum samples taken from CF patients. However, due to the non-specific nature of dityrosine formation [23] no statistically significant correlation was found between its detection and the presence of exclusively neutrophils or eosinophils. The analysis focused upon oxidatively modified tyrosine residues, which produced statistically significant correlations with cellular populations.

Figure V-4 shows a summary of the differences in oxidative metabolites measured by NMR in control and CF populations. A highly significant difference in 3-chlorotyrosine formation was observed between samples from control and the CF populations ($p < 0.01$, Figure V-4a). By incorporating sputum neutrophil cell counts from CF patient sputum samples I was able to standardize the amount of 3-chlorotyrosine produced by a million neutrophils in CF patients' sputum. Within the CF samples, the standardized 3-chlorotyrosine production averaged 5.5 mM ± 0.5 mM 3-chlorotyrosine/10^6 neutrophils. Although this is higher than that obtained with neutrophils stimulated *in vitro*, the finding is in agreement with earlier papers investigating 3-chlorotyrosine production taken from CF sputum [24, 25] (5.82 mM ± 0.76 mM) and suggests that neutrophil activation *in vivo* exceeds that achieved under *in vitro* conditions.

Interestingly, brominated tyrosine in CF sputum was found to be at levels that were elevated and statistically significant when compared to control levels (Figure V-4b, $p < 0.05$ 3-bromotyrosine, and Figure V-4c, $p < 0.01$ 3,5-dibromotyrosine). Standardization of 3-bromotyrosine production to eosinophil cell counts was 4.94 mM ± 0.51 mM 3-bromotyrosine/10^6 eosinophils in CF sputum, which is higher than values obtained *in vitro*. Despite the finding that eosinophil cell counts were similar in control and CF subjects, NMR analysis showed elevated levels of brominated tyrosine residues in CF sputum. This finding was exciting as it showed that eosinophil activation occurs in the CF airway.

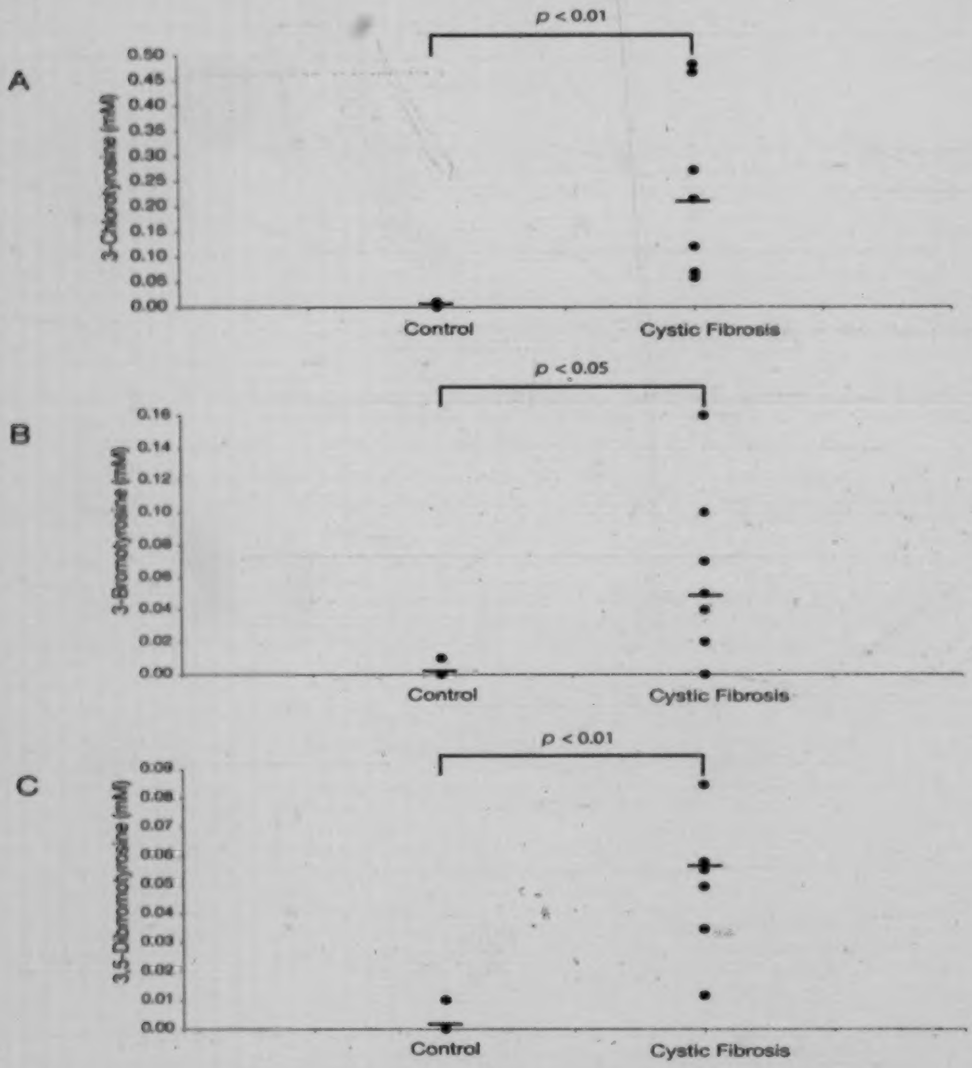


Figure V-4. **Oxidized tyrosine formation in CF and normal subjects**. Oxidative modification of free tyrosine residues in control and CF patient populations as measured by NMR. (A) 3-chlorotyrosine, (B) 3-bromotyrosine, and (C) 3,5-dibromotyrosine levels are shown for control and CF populations. Horizontal bars represent population median.

There were also several significant correlations observed between cell counts and modified tyrosine residue formation in sputum samples from CF patients, as described in the correlation matrix (Table V-2). A statistically significant relationships was observed between 3-chlorotyrosine and neutrophil cell counts ($r^2 = 0.869$, $p = 0.01$). In addition, a strongly negative correlation was found for 3-chlorotyrosine and macrophage counts ($r^2 = -0.838$, $p < 0.02$) in samples from CF patients (Table V-2). This suggests that macrophages, which also produce MPO and reactive oxygen species in a similar manner to neutrophils, do not significantly contribute to the production of 3-chlorotyrosine in the CF airway. The inverse correlation of increasing neutrophil populations with decreasing eosinophil, macrophage, and lymphocyte populations may also indicate the unique pathophysiology and immunological response indicative of CF patients.

	Neutrophil	Eosinophil	Macrophage	Lymphocyte	3-Bromotyrosine	3-Chorotyrosine	Dibromotyrosine
Neutrophil	1						
Eosinophil	-0.83	1					
Macrophage	-0.95	0.68	1				
Lymphocyte	-0.89	0.80	0.79	1			
3-Bromotyrosine	0.43	-0.04	-0.39	0.39	1		
3-Chlorotyrosine	0.87	-0.56	-0.84	-0.76	0.61	1	
Dibromotyrosine	0.33	0.17	-0.34	-0.29	0.83	0.60	1

Table V-2. Bivariate correlation of CF patient oxidative metabolites

a. Bivariate correlations between cell counts and indices of oxidative metabolites from sputum samples of CF patients as measured with high-resolution ^{1}H-NMR

A regression analysis was performed on the CF data showing the linear relationship between clinical cell counts and modified tyrosine residues as measured by NMR (Figure V-5). A positive correlation was found when plotting neutrophil populations and 3-chlorotyrosine production (Figure V-5A); however, negative correlations were found when plotting macrophage or lymphocyte populations and 3-chlorotyrosine (Figure V-5B, C). These findings suggest that the neutrophil is a primary source of 3-chlorotyrosine formation in the airways of CF patients.

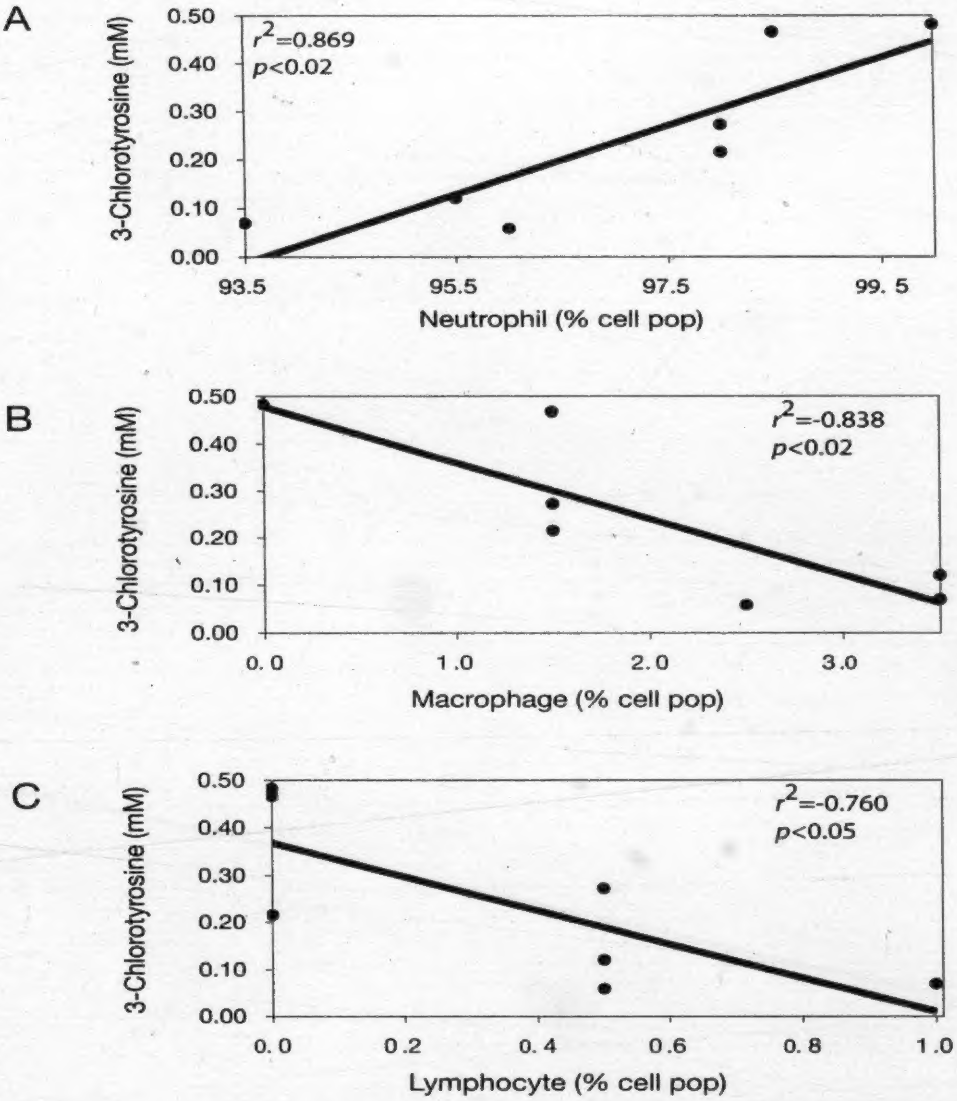


Figure V-5. **Correlations between cellular infiltration and 3-chlorotyrosine**. Graphical representation of the relationship between clinical cell counts and indices of oxidative damage, as measured with NMR, within CF patients. Graphical correlations are demonstrated between (A) 3-chlorotyrosine and neutrophils, (B) 3-chlorotyrosine and macrophages, and (C) 3-chlorotyrosine and lymphocyte cell counts.

DISCUSSION

This study reports the detection of modified tyrosine residues by NMR spectroscopy in purified cell preparations of eosinophils and neutrophils, and confirms their presence in sputum samples from control subjects and patients with CF. This study has demonstrated that neutrophils produce chlorinated tyrosine residues during stimulation by CB/fMLP, a potent secretagogue and inducer of respiratory burst in these cells, as previously shown[26]. The amount of 3-chlorotyrosine detected following *in vitro* stimulation of peripheral blood neutrophils was slightly elevated when compared with previously published results [24](4.5 ± 1.0 μM 3-chlorotyrosine, Eiserich et al.). This may be explained by the focus of Eiserich et al. on the modification of peptidyl tyrosine residues, which showed a significant dependence upon the primary structure, as compared to my analysis of free tyrosine residues that are easily modified following *in vitro* neutrophil activation. Similarly, I found that human peripheral blood eosinophils generated brominated tyrosine residues 3-bromotyrosine and 3,5-dibromotyrosine following stimulation with a cytokine cocktail (IL-3, IL-5, and GM-CSF).

The formation of modified tyrosine residues has been previously demonstrated following exposure to chemical brominating agents (e.g. HOBr, N-bromo amines, and N, N-dibromoamines) using a cell-free system, or by stimulating peripheral blood eosinophils[13, 14]. However, samples were typically analyzed by mass spectrometry (MS) or high performance liquid chromatography (HPLC) [26, 27]. This study is the first to use high-resolution ^{1}H

NMR spectroscopy to identify modified tyrosine residues generated by stimulated peripheral blood eosinophils and neutrophils, as well as in induced sputum samples from control subjects and CF patients. This is of great importance given the increased ease with which NMR identification can be carried out. HPLC and MS require the addition of numerous chemicals during sample preparation and data acquisition, which may inadvertently modify the final results. NMR requires little to no sample preparation and takes only a few minutes to collect the data. The samples are not modified in any way and may be frozen and re-analyzed at a later date. The ease of data acquisition and sample preparation, in conjunction with findings in agreement with earlier studies regarding CF and cellular stimulation accommodates the hypothesis that NMR can serve as a tool for *in vitro* and *in vivo* studies.

In CF sputum, neutrophils were substantially elevated and their numbers correlated strongly with the presence of 3-chlorotyrosine, confirming an earlier study, which used gas chromatography-mass spectrometry to detect 3-chlorotyrosine in CF sputum[28]. The formation of chlorinated tyrosine residues is dependent on MPO-mediated catalysis of physiologically relevant concentrations of bromide ions and H_2O_2, using free L-tyrosine as a substrate for the reaction[24, 25, 29]. There was a negative correlation between 3-chlorotyrosine formation and macrophage counts, suggesting that 3-chlorotyrosine formation is specific to neutrophilic infiltration. The strong correlation between neutrophil counts and 3-chlorotyrosine levels in sputum samples from CF patients also suggests an association between airway neutrophil activation and potential oxidative damage.

The analysis of *in vitro* stimulation of eosinophils in this study demonstrated similar 3-bromotyrosine production levels to those described in previous publications[13]. Eosinophils were detected in CF sputum and the number of infiltrated eosinophils were correlated to 3-bromotyrosine and 3,5-dibromotyrosine formation using NMR metabolite quantitation of mucus plugs from sputum. Because of the relatively small eosinophil population in CF sputum, cellular correlations with tyrosine residue oxidative modification were weakened. However, brominated tyrosine residues were detectable at elevated levels in the sputum of CF patients. These findings suggest that eosinophils, although present in numbers similar to those of control subjects, may well be activated in the airways of CF patients. Intact sputum eosinophils in CF airways may degranulate *in situ* as well as release superoxide (which is rapidly dismutated to form H_2O_2) through the activation of a respiratory burst. However, due to the complex nature of substrate specificity of peroxidases, which occasionally show reactivity with other halides, I can only suggest that 3-bromotyrosine and 3,5-dibromotyrosine may be specific markers of eosinophil activation in tissues and airways.

When normalized for cell counts, the amount of modified tyrosine production in CF sputum suggests a slightly heightened degree of activation and degranulation of both eosinophils and neutrophils in sputum compared to peripheral blood cells from control subjects during *in vitro* stimulation. The increased degree of neutrophil and eosinophil activation in CF sputum samples is suggestive of multiple pathways of cellular stimulation in CF airways. This finding reinforces the need for a greater understanding of how these cells, particularly eosinophils, behave in CF.

In summary, high-resolution NMR spectroscopic analysis has been shown to be a rapid and efficient approach for the measurement of metabolites in sputum samples in patients with CF, and may lend itself readily to metabolomic analysis of disease phenotypes. In addition to the unique findings of this study, much of the preliminary data is in agreement with previous studies, which utilized HPLC or MS for data analysis, thereby strengthening and ensuring the analytical accuracy of high-resolution NMR and the application to *in vivo* studies. The advantage of high-resolution NMR over other analytical methods such as HPLC, or GC is evident in the ability of NMR to detect all the constituents of a sample, which includes a wide range of proteins, carbohydrates, and metabolites, without sample purification, processing, or modification. By using this technique this study was able to demonstrate a strong correlation between neutrophil counts and 3-chlorotyrosine formation in CF sputum. In addition, I was surprised to detect the presence of eosinophil-derived 3-bromotyrosine in sputum samples from CF patients. Although this study focused on the production of modified tyrosine residues in sputum samples, NMR may be an important new tool in the search for broader diagnostic and prognostic applications.

LITERATURE CITED

1. Devidas, S. and W.B. Guggino, The cystic fibrosis transmembrane conductance regulator and ATP. Curr Opin Cell Biol, 1997. 9(4); 547-52.
2. Elborn, J.S. and D.J. Shale, Cystic fibrosis. 2. Lung injury in cystic fibrosis. Thorax, 1990. 45(12); 970-3.
3. Rosenstein, B.J. and P.L. Zeitlin, Cystic fibrosis. Lancet, 1998. 351(9098); 277-82.
4. Danel, C., et al., Quantitative assessment of the epithelial and inflammatory cell populations in large airways of normals and individuals with cystic fibrosis. Am J Respir Crit Care Med, 1996. 153(1); 362-8.
5. Harrison, J.E. and J. Schultz, Studies on the chlorinating activity of myeloperoxidase. J Biol Chem, 1976. 251(5); 1371-4.
6. Babior, B.M., NADPH oxidase: an update. Blood, 1999. 93(5); 1464-76.
7. Klebanoff, S.J., Oxygen metabolism and the toxic properties of phagocytes. Ann Intern Med, 1980. 93(3); 480-9.
8. Mohammed, J., et al., Purification and cytotoxic potential of myeloperoxidase in cystic fibrosis sputum. J. Lab Clin. Med., 1998. 112(6); 11-20.
9. Fu, S., et al., Reactions of hypochlorous acid with tyrosine and peptidyl-tyrosyl residues give dichlorinated and aldehydic products in addition to 3-chlorotyrosine. J Biol Chem, 2000. 275(15); 10851-8.
10. Kita, H., C. Adolphson, and G. Gleich, Biology of eosinophils, in Allergy: principles and practice, N.J. Adkinson, et al., Editors. 1998, Mosby: St. Louis. p. 242-50.
11. Lacy, P., et al., Divergence of mechanisms regulating respiratory burst in blood and sputum eosinophils and neutrophils from atopic subjects. J Immunol, 2003. 170(5); 2670-9.
12. Lacy, P. and R. Moqbel, Immune effector functions of eosinophils in allergic airway inflammation. Curr Opin Allergy Clin Immunol, 2001. 1(1); 79-84.

13. Wu, W., et al., Eosinophils generate brominating oxidants in allergen-induced asthma. J Clin Invest, 2000. 105(10); 1455-63.

14. Wu, W., et al., 3-Bromotyrosine and 3,5-dibromotyrosine are major products of protein oxidation by eosinophil peroxidase: potential markers for eosinophil-dependent tissue injury in vivo. Biochemistry, 1999. 38(12); 3538-48.

15. Pizzichini, M.M., et al., Sputum in severe exacerbations of asthma: kinetics of inflammatory indices after prednisone treatment. Am J Respir Crit Care Med, 1997. 155(5); 1501-8.

16. Pizzichini, M.M., et al., Spontaneous and induced sputum to measure indices of airway inflammation in asthma. Am J Respir Crit Care Med, 1996. 154(4 Pt 1); 866-9.

17. Gauvreau, G.M., et al., Effects of inhaled budesonide on allergen-induced airway responses and airway inflammation. Am J Respir Crit Care Med, 1996. 154(5); 1267-71.

18. Pizzichini, E., et al., Sputum eosinophilia predicts benefit from prednisone in smokers with chronic obstructive bronchitis. Am J Respir Crit Care Med, 1998. 158(5 Pt 1); 1511-7.

19. Levi-Schaffer, F., et al., Association of granulocyte-macrophage colony-stimulating factor with the crystalloid granules of human eosinophils. Blood, 1995. 85(9); 2579-86.

20. Lacy, P., et al., Rapid mobilization of intracellularly stored RANTES in response to interferon-gamma in human eosinophils. Blood, 1999. 94(1); 23-32.

21. Wei, J., et al., Low concentrations of serum tyrosine in neuroleptic-free schizophrenics with an early onset. Schizophr Res, 1995. 14(3); 257-60.

22. Reinhardt, N., et al., Cellular profiles of induced sputum in children with stable cystic fibrosis: comparison with BAL. Eur Respir J, 2003. 22(3); 497-502.

23. Giulivi, C., N.J. Traaseth, and K.J. Davies, Tyrosine oxidation products: analysis and biological relevance. Amino Acids, 2003. 25(3-4); 227-32.

24. Eiserich, J.P., et al., Formation of nitric oxide-derived inflammatory oxidants by myeloperoxidase in neutrophils. Nature, 1998. 391(6665); 393-7.

25. Van Der Vliet, A., et al., Myeloperoxidase and protein oxidation in cystic fibrosis. Am J Physiol Lung Cell Mol Physiol, 2000. 279(3); L537-46.

26. Coxon, P.Y., et al., MAPK-activated protein kinase-2 participates in p38 MAPK-dependent and ERK-dependent functions in human neutrophils. Cell Signal, 2003. 15(11); 993-1001.

27. Schneider, T. and A.C. Issekutz, Quantitation of eosinophil and neutrophil infiltration into rat lung by specific assays for eosinophil peroxidase and myeloperoxidase. Application in a Brown Norway rat model of allergic pulmonary inflammation. J Immunol Methods, 1996. 198(1); 1-14.

28. van der Vliet, A., et al., Analysis of aromatic nitration, chlorination, and hydroxylation by gas chromatography-mass spectrometry. Methods Enzymol, 1999. 301; 471-83.

29. Crow, J.P., Measurement and significance of free and protein-bound 3-nitrotyrosine, 3-chlorotyrosine, and free 3-nitro-4-hydroxyphenylacetic acid in biologic samples: a high-performance liquid chromatography method using electrochemical detection. Methods Enzymol, 1999. 301; 151-60.

CHAPTER VI

Animal model of asthma

OVERVIEW

As an asthma attack develops immunological effector cells are sequestered to the lungs and initiate an inflammatory response. Current pharmacological measures, such as corticosteroids, attempt to control the severity of inflammation in an effort to maintain proper airflow in the lungs. Typically, measurements of pulmonary inflammation rely on invasive procedures, such as induced sputum or bronchoscopy, to determine the severity and physiological process. *This chapter highlights a study into the ability of NMR to detect on-going inflammation in the lungs of a guinea pig model of asthma.* Urine was collected and analyzed for key metabolites, in an effort to obtain a biochemical understanding of the immunological effector mechanisms, access the progression of inflammation, effects on pulmonary function, and overall pathophysiology. The ability of NMR to detect indices of changing pulmonary physiological states, in a patient's urine, during an asthma attack could provide for a non-invasive technique to monitor pulmonary inflammation and direct therapeutic treatment in future human studies.

INTRODUCTION

Asthma exacerbations are responsible for over 1.9 million emergency department visits in America[1], and roughly 500,000 hospitalizations annually[1]. The worldwide market for asthma medication is valued at $5.5 billion a year (U.S. dollars) [2]. A national survey in Canada reported that approximately 60% of individuals with asthma have poor control of the disease as evident in short-term symptoms, medication use, functional impact, and long-term disease burden[3]. For an individual patient with acute exacerbation of asthma, an emergency department visit will cost an estimated $324 and over $600 per day if hospitalized (CDN dollars) [4]. The growing population affected by asthma and the cost of treatment demonstrates the need for a non-invasive method of diagnosis, and the monitoring of therapeutics.

Traditionally, asthma was thought of as a disease dictated by airway smooth muscle contraction and airflow obstruction[5]. Treatment typically rested upon the use of bronchodilators. It is now understood that asthma is a disease manifested by inflammation and as such is optimally managed through treatment with anti-inflammatory medications (*e.g.* inhaled corticosteroids) [6-10]. The inflammatory response in asthma is characterized by the activity of many different immunological cells, such as mast cells, eosinophils, basophils, and lymphocytes; as well as their secreted products [2, 11-21]. Together these cells and their effector mechanisms produce an altered airway, which exhibits microvascular permeability with exudation of plasma, increased mucus secretion, desquamation of epithelial cells, airway hyperresponsiveness, and structural remodeling such as peribronchial collagen deposition and hypertrophy of airway smooth muscle[2, 16-21].

The severity of asthma and the degree of airway hyperresponsiveness correlates with the degree of inflammation [22]. However, the mechanism through which inflammation causes airway hyperresponsiveness and airway obstruction is largely unknown[23]. Data suggest that the mediators released from the effector cells in the airway cause airway damage at many levels and subsequent activation of additional inflammatory cells. Release of mediators such as leukotrienes, histamine, tryptase, and eosinophil granule products (*i.e.* EPO, major basic protein (MBP), and eosinophil derived neurotoxin (EDN)) cause bronchoconstriction, airway hyperresponsiveness, and the clinical manifestations of inflammation (*e.g.* cough, wheeze, and dyspnea)[12, 24].

One methodological block to the elucidation of many of the mechanisms implicit to human asthma is the difficulty of measuring inflammatory cells and their mediators qualitatively and quantitatively in an accurate manner. The ability to detect and monitor airway inflammation accurately is key to understanding and managing the pathophysiology associated with asthma[12, 25]. The early detection and measurement of inflammation could provide information about diseased airways that may not be revealed by other techniques presently used in clinics (*e.g.* symptoms, pulmonary function, and airway hyper-responsiveness)[12, 25]. For example, indicators of asthma such as symptoms or pulmonary function tests represent the physiological endpoint of asthma, limited airflow. A more appropriate clinical management of asthma would involve the identification of an impending acute episode to ensure the disease is controlled as much as possible. A metabolomic study of asthma could allow for the idenitification of on-going physiological changes in the airways and might allow for improved management of the disease.

An appropriate preliminary study of asthma urine metabolites would involve the correlation of pulmonary inflammatory mechanisms with different urine metabolites detected by NMR, and accepted laboratory techniques (*e.g.* differential cell counts and lung function). The animal model of asthma provides an opportunity to control the disease process of asthma, perform pulmonary function analysis, and pulmonary cellular analysis, as well as limit the degree of environmental influences that often confound human studies. The guinea pig is frequently used as an animal model because of the similarity of airway physiology to the human, as well as the functional analysis of cellular and mediator responses. Guinea pig models of bronchopulmonary inflammation have shown to be extremely useful when examining basic mechanisms of allergic inflammation and the underlying physiological response. Guinea pig models have helped to create an understanding of the roles of airway smooth muscle[26], the mediators and adhesion molecules involved in leukocyte recruitment[27, 28], goblet cell hyperplasia[29], the role of eosinophils in the allergic response[30, 31], and neuronal controls[32-34].

In collaboration with Dr. Adamko,. UofA, a guinea pig model of asthma was produced that demonstrated cellular infiltration and pulmonary function in a manner similar to human asthma pathophysiology. Urine was collected from control, sensitized, and challenged guinea pigs, and analyzed by NMR. My sincere thanks to Dr. Idongesit Obiefuna who performed all of the animal and histological work. Qualitative and quantitative spectral analysis, as well as the use of multivariate statistical tools identified spectral regions and metabolites that were unique to the different physiological states of the animals

EXPERIMENTAL PROCEDURES

VI-A. Guinea Pig Animals

Specific pathogen-free female Dunkin-Hartley guinea pigs (180–450 g; Charles River Laboratories) were used. All animals were shipped in filtered crates and kept in high-efficiency particulate- filtered air, and fed a normal guinea pig diet (Prolab; Agway). The animals were handled in accordance with the standards established by the Health Sciences Animal Policy and Welfare Committee, guidelines of the Canadian Council on Animal Care, and requirements of the provincial legislation entitled 'The Universities Act", Section 52, of the province of Alberta.

VI-B. Guinea Pig Treatment

Some animals were sensitized to ovalbumin (OVA) (Sigma-Aldrich, Ont.) by intraperitoneal injection with 10 mg/kg OVA (0.3 ml) on days 0 and 3. Twenty-one days later some of the animals were exposed to nebulized 0.5% ovalbumin by inhalation for 10–20s (challenged group). Studies of airway function, inflammatory response, and lung histological studies were performed four days after ovalbumin challenge. Nonsensitized, pathogen-free guinea pigs were used as controls. Some control animals were subjected to inhalation of ovalbumin without prior sensitization to confirm that inhalation alone did not induce airway pathology (data not shown). The final number of animals included in the study were; 24 control, 23 sensitized, and 29 challenged guinea pigs.

VI-C. Airway responsiveness to histamine

Experiments were conducted four days after antigen challenge. The guinea pigs were anesthetized intraperitoneally with urethane (1.5 g/kg) (Sigma-Aldrich, Ont.). This dose produced a deep anesthesia lasting 8–10 hours although none of these experiments lasted longer than four hours (Green, 1982). Adequate depth of anesthesia was assessed via the pedal reflex. A cannula was placed in the left jugular vein to allow for the administration of histamine solutions. The trachea was cannulated and the animals were ventilated with a positive pressure, constant volume rodent respirator (Harvard Apparatus, Inc., MA) at a tidal volume of 10 ml/kg and a respiratory rate of 100 breaths/min. The animals were paralyzed by intravenously infusing succinylcholine (10 μg/kg/min). The animal's body temperature was maintained at 37°C using a water heating pad. Pulmonary inflation pressure (Ppi) was measured at the trachea using a DTX™ pressure transducer. All signals were recorded by AD Instruments software (Mountain View, CA). Bronchoconstriction was measured as the increase in Ppi above the basal inflation pressure produced by the ventilator. With this method, increases in Ppi as small as 2–3 mm H_2O could be accurately recorded. Increasing bolus doses of histamine (1–50 μg·kg-1 i.v.) were administered at 6 min. intervals and the resulting bronchoconstrictor responses were recorded as increases in PPi.

VI-D. Assessment of Pulmonary Inflammation – Bronchoalveolar Lavage

At the end of the experiment, bronchoalveolar lavage was performed via the tracheal cannula. The lungs were lavaged with 5 aliquots of 10.0 ml phosphate buffered saline (PBS). The recovered lavage fluid (40–45 ml) was centrifuged (350 g for 7min). The cells were resuspended briefly in 10 ml of deionized water to remove any erythrocytes before an additional 40 ml of PBS was added. Cells were centrifuged again, the supernatant was poured off, and the cells were resuspended in 10 ml of PBS. Cells were counted using a Neubauer Hemocytometer (Hausser Scientific Co., PA). Aliquots of the cell suspension were spun onto glass slides, stained with Diff-Quik® (Baxter Healthcare Corp., Deerfield, IL), and counted to obtain differential cell counts. At least 200 cells were counted on each slide. Cells were differentiated using standard morphological criteria.

VI-E. Histological Evaluation

At the end of each experiment, the lungs were removed, inflated and fixed with 3.7% formaldehyde in PBS. Twenty-four hours later the lungs were removed from fixative and transverse sections of the trachea and each lobe of the lung were taken for histological evaluation. Sections were embedded in paraffin blocks, sliced in consecutive 6.0 μm sections, and mounted onto glass slides for light microscopy.

VI-F. Histological Identification of Eosinophils

A carbol chromotrope stain was used to identify guinea pig eosinophils, with an adaptation of the method described by Lendrum[35]. This stain was concentrated in the granules of eosinophils due to their avidity for acid dyes and has been shown by several investigators to be specific for eosinophils. Briefly, paraffin slides were dewaxed in xylenes, rehydrated through graded alcohols, and washed in PBS. The slides were transferred to Mayer's hematoxylin solution for 4 min and then washed in deionized water. Next, they were incubated in acidified alcohol (1% HCL, 70% EtOH) for 2 min, washed in deionized water, incubated in an aqueous solution of 1% wt/vol chromotrope 2R (Sigma, St. Louis, MO) with 5% wt/vol phenol (Sigma) for 20 min, and washed again in deionized water. The slides were then dehydrated and a cover slip was applied.

VI-G. Histological Analysis

Two cartilaginous bronchi from each of the animal groups were examined. Airways were selected by starting at the top left corner of the slide and moving in a counter clockwise manner. An obvious landmark within an airway was chosen under low power as the starting point for analysis. Beginning at this point, the sections were viewed under an oil immersion lens (100X). The image was oriented so that the basement membrane was uppermost, thus, a section of airway containing submucosa and airway smooth muscle was examined. Ten consecutive sections were examined and eosinophil numbers per high power field was determined.

VI-H. Guinea Pig Urine Collection

While under anaesthesia urine samples were collected by trans-abdominal cystocentesis using a sterile 21-gauge needle. The needle was inserted at a 45° angle, on the midline, midway between the umbilicus and brim of the pelvis, while creating negative pressure by pulling back on the plunger of the syringe. If urine was not obtained with the first puncture, two additional punctures were attempted from 0.2 to 1.0 centimeters cranial or caudal to the initial puncture site. The needle was changed before making each attempt to avoid contamination. Urine samples (1.0-2.0 cc) were stored at -80°C for later ^{1}H-NMR analysis.

VI-I. Urine Sample Preparation

Samples were catalogued and stored at the Canadian National High Field NMR Centre (NANUC) at -80°C. Urine samples were thawed in a biosafety fume hood and a 630 μl aliquot was removed and placed in a 1.5 ml Eppendorff tube followed by the addition of 70 μl of a reference buffer solution ((4.9 mM DSS (disodium-2, 2-dimethyl 2-silapentane-5-sulphonate) and 100 mM imidazole in D_2O) Sigma-Aldrich, Mississauga, ON). Each sample was then brought to a pH of 7.0 +/- 0.1 using HCl and NaOH. An aliquot of 600 μl was taken and transferred to a standard 5 mm glass NMR tube (Wilmad, NJ).

VI-J. NMR Analysis

All ^{1}H-NMR spectra were acquired on a 600 MHz Inova (Varian Inc, Palo Alto, CA.) spectrometer equipped with a 5 mm triple-resonance (HCN) probe with z-axis gradient coil. One-dimensional ^{1}H-NMR spectra were collected at 25°C with a tnnoesy pulse sequence (one-dimensional, three pulse NOESY, with a transmitter pre-saturation delay of 900 ms for water suppression during the pre-acquisition delay and 100 ms mixing time) and a spectral width of 7200 Hz. The time-domain data points were 64k complex points, acquisition time was 4s, the 90° pulse was 6.8μs, repetition time was 5s, 4 steady state scans, and the number of acquired scans was 32 per FID. The data was apodized with an exponential window function corresponding to a line broadening of 0.5 Hz, zero-filled to 128k complex points, and Fourier transformed [36].

VI-K. NMR Quantification

The methyl singlet of the buffer constituent DSS served as internal standard for chemical shifts (set to 0 ppm) and for quantification. Spectral identification and quantification of 24 clearly identifiable metabolites were performed using the Chenomx NMR Suite Professional software package Version 3.1 (database available at pH 7.0, Chenomx Inc., Edmonton, AB). The internal DSS signal was utilized as the concentration reference (0.49 mM). The software included an appropriate correction for urine dilution due to the addition of buffer. Work in the laboratory demonstrated that this procedure provides absolute concentration accuracies in excess of 90%[37].

VI-L. Multivariate Statistical Analysis

The xytrace of the NMR spectra (ASCII format) were sent to Brion Dolenko and Ray Somorjai at the Institute for Biodiagnostics (Winnipeg, MB.) for multivariate analysis.

The guinea pig data was comprised of three classes: Normal (24), Sensitized (ovalbumin sensitized), and Challenged (ovalbumin sensitized + ovalbumin challenge) (29) (23). Three pair classifiers were developed; Normal vs. Challenged, Sensitized vs. Challenged, and Normal vs. Sensitized. For all three pair classifiers, the optimal spectral regions were first determined. This feature selection was carried out by a genetic-algorithm-based feature selection approach[38]. The feature selection was wrapper-based: *i.e.* a classifier, in this case linear discriminate analysis (LDA), with leave-one-out (LOO) internal cross-validation was used to decide which were the best features.

Having obtained the optimal feature set for each pair of classifier, external cross validation (EXV) helped estimate a more realistic error estimate. EXV is carried out by splitting randomly each dataset into a training set and a test set (50:50). The splitting is stratified, *i.e.*, the relative proportions of the samples in the two classes are retained in the splits. Ten splits were done for each dataset and the averages and standard deviations calculated. Given the small sample sizes, this gives more realistic results than using the entire dataset. Details of the statistical classification techniques used are found in a recent textbook[39].

VI-M. Statistical Analysis

All guinea pig physiological data were expressed as mean and standard error of the mean (SEM). Baseline heart rates, blood pressures, Ppi, histological measurements, and bronchoalveolar lavage were analyzed using analysis of variance (Statview 4.5; Abacus Concepts, Inc.). A P-value of $\leq$ 0.05 was considered significant.

RESULTS

VI-A. Pulmonary Function

There were no significant differences between the groups for animal weight (kg) (control, 0.449 ± 0.02; sensitized, 0.476 ± 0.03; sensitized challenged 0.434 ± 0.02). A positive pressure of 60-170 mmH_2O (mean 103± 3.4 mmH_20) was needed to ventilate the animals. Sensitization of pathogen free guinea pigs did not alter baseline pulmonary inflation pressure (80.0±3.1, n=25) compared to nonsensitized controls (76.0±3.4, n=25). Challenge with the allergen plus sensitization, increased baseline pulmonary inflation pressure compared to respective controls (130.0±6.0, n=25).

VI-B. Ovalbumin Sensitization and Challenge Induces Airway Hyperreactivity

In the guinea pigs, histamine (1-20μg kg^{-1} i.v.) caused a dose-dependent bronchoconstriction in the control animals (black crosshairs, Figure VI-1). This was not affected by sensitization as dose-dependent bronchoconstriction was similar to control animals (blue boxes, Figure VI-1) In contrast, ovalbumin sensitization and challenge lead to airway hyperreactivity as the degree of bronchoconstriction increased compared to control (red triangles, Figure VI-1).

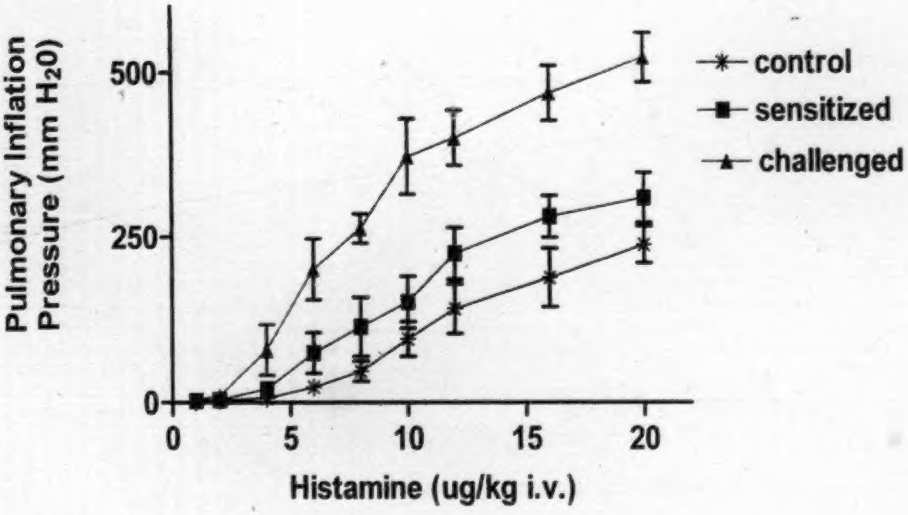


Figure VI-1. **Antigen challenged animals develop airway hyperresponsiveness**. Control and sensitized animals produced a similar response to histamine exposure (n=25). Allergen sensitized and challenged guinea pigs produced a marked response histamine exposure (n=25), indicating airway hyperreactivity.

VI-C. Ovalbumin Sensitization and Challenge Induces Airway Inflammation

At the end of each experiment, lung lavage was performed to quantitate the number of inflammatory cells in the lumen of the lung. The total number of cells recovered by lung lavage was increased by allergen sensitization and challenge (Figure VI-2). The cellular increase was comprised of macrophages, eosinophils and neutrophils. Lymphocyte numbers were not significantly affected by the challenge. Sensitization significantly increased the number of eosinophils recovered in the lavage fluid when compared to nonsensitized controls.

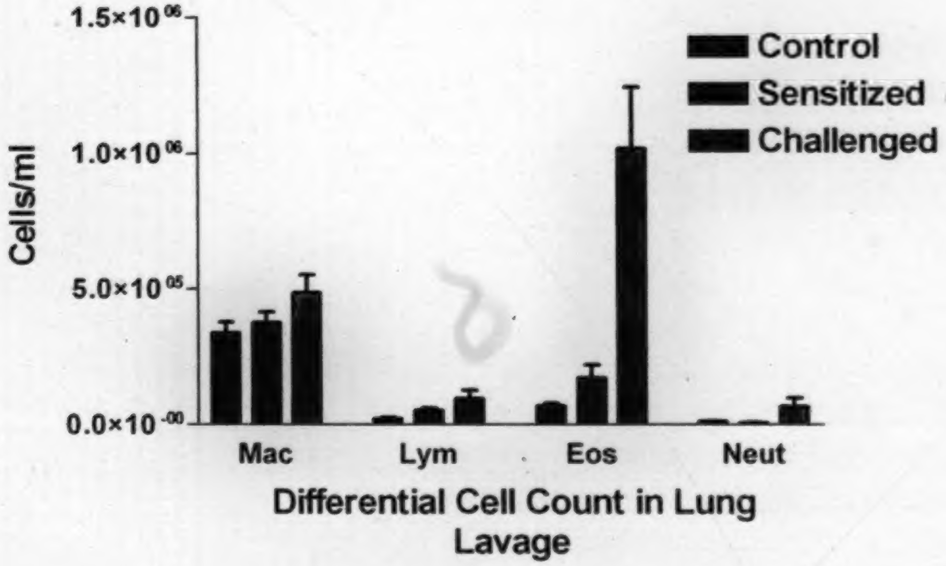


Figure VI-2. **Antigen challenge induces inflammatory cell influx to airway lumen**. Total cell counts were significantly increased in sensitized as well as sensitized-challenged guinea pigs. Lymphocytes were not significantly affected by challenge. Lavage cell populations were comprised of macrophages (Mac), lymphocytes (Lym), eosinophils (Eos), and neutrophils (Neut).

VI-D. Airway Tissue Inflammation

Airway sections were stained to quantitate the average number of eosinophils per high power field over the entire thickness of the airway (Figure VI-3). Sensitization alone caused increased presence of eosinophils in the airway wall. Challenge of sensitized animals caused a further increase in eosinophil number.

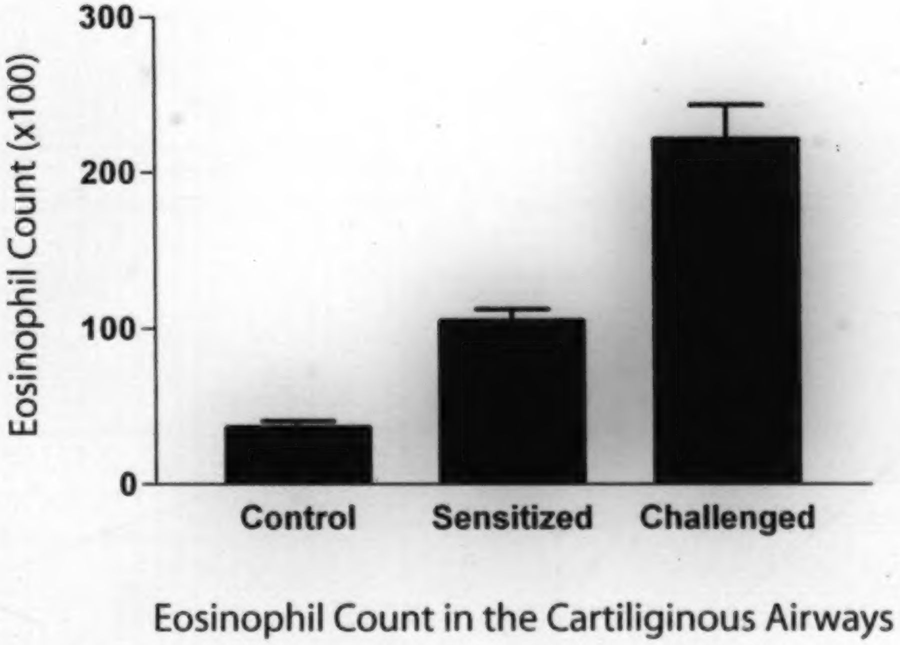


Figure VI-3. **Antigen challenge induces eosinophil influx into airway walls.** Sensitization caused an increase in the number of eosinophils in the wall of the airway. Challenged guinea pigs produced a greater influx of eosinophils.

VI-E. Multivariate Statistical Analysis

Multivariate analysis of the guinea pig urine returned three NMR spectral regions that were able to separate the different populations. The ability of the multivariate analysis to separate the different populations was expressed as an average accuracy, which reports the probability that the software correctly separated the populations over three different attempts, see Table VI-1.

Guinea Pig Treatment	Average Accuracy (%)
Normal vs. Sensitized	87.1 ± 8.7
Normal vs. Challenged	82.5 ± 11
Sensitized vs. Challenged	85.2 ± 13

Table VI-1. Accuracy of guinea pig diagnosis by multivariate analysis

a. Likelihood that the multivariate analysis will correctly separate urine based upon NMR spectra

b. Expressed as average accuracy over three attempts (± standard deviation)

VI-F. NMR and Biochemical Interpretation

Multivariate analysis returned the spectral regions and signatures that were able to differentiate the different guinea pig populations. Translating the multivariate information to the original NMR spectra allowed for the identification of the spectral resonance peaks and the metabolites that separated the different guinea pigs.

DISCUSSION

Airway hyperreactivity, marked increase of inflammation in the lumen and walls of the airway, and obstructed airflow are classic indicators of asthma. Asthma places large amounts of stress on the human body, there are continuing physiological changes in the airway, as well as an increase in immunological activity in lung. Clinically, there remains great difficulty in the diagnosis of asthma, in particular, the diagnosis of the very young. It was hypothesized that urine could be sampled from a guinea pig model of asthma and analyzed by NMR spectroscopy. This metabolomic analysis was hypothesized to identify key metabolites in the urine that correlate with the diseased state of the animal. The development of NMR as a non-invasive tool to identify metabolites indicative of asthma would aid in the clinical identification and intervention of asthma, as well as the current understanding of asthma pathophysiology.

Pulmonary function analysis, as assessed by the pulmonary inflation pressure, revealed that control and sensitized animals had similar function. Sensitization followed by an additional challenge with the allergen OVA caused a significant increase in the airway reactivity to histamine. This indicated that the challenged guinea pig produced a hyperreactive and bronchoconstrictive response to allergen exposure (Figure VI-1). These types of responses by the challenged guinea pig are in agreement with the current understanding of airway hyperreactivity in humans.

The cellular infiltrates recovered by BAL indicate that the challenged guinea pig underwent active recruitment of immunological cells to the lumen of the lung. Macrophages increased slightly for each of the control, sensitized, and challenged, but the number lumen eosinophils in the challenged guinea pig increased significantly, indicating an active immunological response (Figure VI-2). Histological examination of sectioned airways also revealed significantly increased eosinophil deposition in the walls of challenged guinea pig airways (Figure VI-3). Although sensitized animals revealed an increase in the presence of eosinophils when compared with controls (roughly twice the amount), the challenged animals produced an eosinophil count ten times that found in the control guinea pigs.

Pulmonary function, hyperreactivity, and histological analysis, along with the differential cell counts revealed that the sensitization and challenge techniques for the animal model were successful in creating a pathophysiology similar to asthma. Urine was sampled from every guinea pig animal and analyzed by 1D ^{1}H-NMR spectroscopy. The spectra were exported for multivariate statistical analysis by LDA and LOO cross-validation. The analysis returned three spectral regions that were able to separate the three guinea pig populations. Time did not allow for the final biochemical identification of the metabolites selected by the algorithms. Using the Chenomx software (Edmonton, AB), which contains a database of over 300 known urine metabolites the NMR signature that was able to separate the different guinea pig populations was not identified. Work is continuing with additional analytical techniques, such as MS, to identify the spectral features indicative of the different guinea pig physiological states.

LITERATURE CITED

1. McCaig, L.F., National Hospital Ambulatory Medical Care Survey: 1998 emergency department summary. Adv Data, 2000(313); 1-23.
2. Sandford, A., T. Weir, and P. Pare, The genetics of asthma. Am J Respir Crit Care Med, 1996. 153(6 Pt 1); 1749-65.
3. Chapman KR, E.P., Grenville A, Dewland P, Zimmerman S, Control of asthma in Canada: failure to achieve guideline targets. Canadian Respiratory Journal, 2001. 8; Suppl. A:35A-40A.
4. Weiss, K.B. and S.D. Sullivan, The health economics of asthma and rhinitis. I. Assessing the economic impact. J Allergy Clin Immunol, 2001. 107(1); 3-8.
5. National Asthma Education and Prevention Program. Expert Panel Report: Guidelines for the Diagnosis and Management of Asthma Update on Selected Topics--2002. J Allergy Clin Immunol, 2002. 110(5 Suppl); S141-219.
6. Russell, G., Very high dose inhaled corticosteroids: panacea or poison? Arch Dis Child, 2006. 91(10); 802-4.
7. Holgate, S.T. and R. Polosa, The mechanisms, diagnosis, and management of severe asthma in adults. Lancet, 2006. 368(9537); 780-93.
8. Cerasoli, F., Jr., Developing the ideal inhaled corticosteroid. Chest, 2006. 130(1 Suppl); 54S-64S.
9. Irwin, R.S. and N.D. Richardson, Side effects with inhaled corticosteroids: the physician's perception. Chest, 2006. 130(1 Suppl); 41S-53S.
10. Eum, S.Y., et al., Inhibition of allergic airways inflammation and airway hyperresponsiveness in mice by dexamethasone: role of eosinophils, IL-5, eotaxin, and IL-13. J Allergy Clin Immunol, 2003. 111(5); 1049-61.
11. Canonica, G.W., Treating asthma as an inflammatory disease. Chest, 2006. 130(1 Suppl); 21S-8S.

12. Cicutto, L.C. and G.P. Downey, Biological markers in diagnosing, monitoring, and treating asthma: a focus on noninvasive measurements. AACN Clin Issues, 2004. 15(1); 97-111.

13. Garlisi, C.G., et al., T cells are necessary for Th2 cytokine production and eosinophil accumulation in airways of antigen-challenged allergic mice. Clin Immunol Immunopathol, 1995. 75(1); 75-83.

14. Gavett, S.H., et al., Depletion of murine CD4+ T lymphocytes prevents antigen-induced airway hyperreactivity and pulmonary eosinophilia. Am J Respir Cell Mol Biol, 1994. 10(6); 587-93.

15. Gonzalo, J.A., et al., Eosinophil recruitment to the lung in a murine model of allergic inflammation. The role of T cells, chemokines, and adhesion receptors. J Clin Invest, 1996. 98(10); 2332-45.

16. Gonzalo, J.A., et al., The coordinated action of CC chemokines in the lung orchestrates allergic inflammation and airway hyperresponsiveness. J Exp Med, 1998. 188(1); 157-67.

17. Hogan, S.P., et al., Interleukin-5-producing CD4+ T cells play a pivotal role in aeroallergen-induced eosinophilia, bronchial hyperreactivity, and lung damage in mice. Am J Respir Crit Care Med, 1998. 157(1); 210-8.

18. Laitinen, L.A., et al., Eosinophilic airway inflammation during exacerbation of asthma and its treatment with inhaled corticosteroid. Am Rev Respir Dis, 1991. 143(2); 423-7.

19. MacLean, J.A., R. Ownbey, and A.D. Luster, T cell-dependent regulation of eotaxin in antigen-induced pulmonary eosinophila. J Exp Med, 1996. 184(4); 1461-9.

20. Mehlhop, P.D., et al., Allergen-induced bronchial hyperreactivity and eosinophilic inflammation occur in the absence of IgE in a mouse model of asthma. Proc Natl Acad Sci U S A, 1997. 94(4); 1344-9.

21. O'Byrne, P., Asthma pathogenesis and allergen-induced late responses. J Allergy Clin Immunol, 1998. 102(5); S85-9.

22. Martin, R.J., et al., Airways inflammation in nocturnal asthma. Am Rev Respir Dis, 1991. 143(2); 351-7.

23. Cockcroft, D.W. and B.E. Davis, Mechanisms of airway hyperresponsiveness. J Allergy Clin Immunol, 2006. 118(3); 551-9; quiz 560-1.

24. Wardlaw, A.J., et al., New insights into the relationship between airway inflammation and asthma. Clin Sci (Lond), 2002. 103(2); 201-11.

25. Green, R.H., et al., Asthma exacerbations and sputum eosinophil counts: a randomised controlled trial. Lancet, 2002. 360(9347); 1715-21.

26. Chitano, P., L. Wang, and T.M. Murphy, Mechanisms of airway smooth muscle relaxation during maturation. Can J Physiol Pharmacol, 2005. 83(10); 833-40.

27. Williams, T.J. and P.J. Jose, Role of eotaxin and related CC chemokines in allergy and asthma. Chem Immunol, 2000. 78; 166-77.

28. Liu, F., et al., Pharmacological characterization of guinea pig chemoattractant receptor-homologous molecule expressed on Th2 cells (CRTH2). Prostaglandins Other Lipid Mediat, 2005. 76(1-4); 133-47.

29. Rogers, D.F., Airway goblet cell hyperplasia in asthma: hypersecretory and anti-inflammatory? Clin Exp Allergy, 2002. 32(8); 1124-7.

30. Smith, N. and F.J. Johnson, Early- and late-phase bronchoconstriction, airway hyper-reactivity and cell influx into the lungs, after 5'-adenosine monophosphate inhalation: comparison with ovalbumin. Clin Exp Allergy, 2005. 35(4); 522-30.

31. Ishiura, Y., et al., In vivo PAF-induced airway eosinophil accumulation reduces bronchial responsiveness to inhaled histamine. Prostaglandins Other Lipid Mediat, 2005. 75(1-4); 1-12.

32. Adamko, D.J., et al., CD8+ T lymphocytes in viral hyperreactivity and M2 muscarinic receptor dysfunction. Am J Respir Crit Care Med, 2003. 167(4); 550-6.

33. Fryer, A.D., et al., Neuronal eotaxin and the effects of CCR3 antagonist on airway hyperreactivity and M2 receptor dysfunction. J Clin Invest, 2006. 116(1); 228-36.

34. Fryer, A.D., et al., Effects of inflammatory cells on neuronal M2 muscarinic receptor function in the lung. Life Sci, 1999. 64(6-7); 449-55.

35. de Cock, P.A., H.J. Tanke, and J.P. Schroder-Van der Elst, The morphology of the bone marrow of the mouse. Acta Histochem, 1980. 67(1); 95-106.

36. Kumar, A., R.R. Ernst, and K. Wuthrich, A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem Biophys Res Commun, 1980. 95(1); 1-6.

37. Saude, E.J., C.M. Slupsky, and B.D. Sykes, Optimization of NMR analysis of biological fluids for quantitative accuracy. Metabolomics, 2006. 2(3); 113-23.

38. Nikulin, A.E., et al., Near-optimal region selection for feature space reduction: novel preprocessing methods for classifying MR spectra. NMR Biomed, 1998. 11(4-5); 209-16.

39. Somorjai, R.L., et al., A data-driven, flexible machine learning strategy for the classification of biomedical data. Artificial intelligence methods and tools for systems biology, Computational Biology Series, ed. W. Dubitzky and F. Azuaje. Vol. 5. 2004: Springer. 67-85.

CHAPTER VII

Analysis of acute and post-treatment human asthma

OVERVIEW

This chapter describes a study where urine was used as the biofluid for the metabolomic investigation of human asthma. Patients that visited the UofA hospital emergency department (ED) with acute asthma provided urine samples during their clinical stay. The patients were prescribed corticosteroids (prednisone), discharged, and returned to the hospital for a two-week follow-up examination where they provided a second urine sample. It was believed that inter-individual metabolite and disease variability might be reduced by repeatedly sampling the same individual (pre/post-treatment). However, given the results from the normal study (Chapter IV) the intra-individual variability may not be reduced by repeated sampling of the same individual. It was anticipated that acute asthma patients would provide the best example of an active asthma pathophysiology allowing for the greatest likelihood of identifying unique metabolites indicative of the disease. Multivariate statistical analysis of 1D ^{1}H-NMR spectra identified patient separation between the pre- and post-prednisone time points. Continuing analysis at the biochemical level identified important metabolites that changed in an asthma patient following two weeks of steroid treatment.

This chapter is intended to provide an update on continuing work in the metabolomic study of asthma. Due to the amount of time required to properly recruit a large enough sample size of acute asthma patients a formal conclusion was not possible, therefore a brief update is provided in this appendix.

INTRODUCTION

The preceding chapters document studies designed to optimize and address important considerations in the application of NMR as an analytical tool of biological samples for clinical metabolomic investigations. The application of NMR metabolomic research for this thesis was to provide a method to easily sample and study human asthma. For the first human study it was considered best to sample exacerbating, or highly inflamed, subjects. Representing patients at the 'extreme end' of the disease acute asthma patients from the University of Alberta Emergency Department were recruited to the study. Urine samples were collected from the participants during their initial hospital visit. It was hypothesized that the progression of the immunological and surrounding physiological activity of severely acute asthma patients would lead to an increased likelihood of capturing various indicative molecules and metabolites in the urine.

EXPERIMENTAL PROCEDURES

VII-A. Acute Asthma Patient Recruitment

During the initial ED visit adult and pediatric patients were diagnosed with an acute asthma exacerbation. The patients provided an initial urine sample, which was stored in an ED freezer (-10°C) for no more than 2 hr before being transferred to either the deep-freeze (-80°C) at the Canadian National High Field NMR Centre (NANUC) or the clinical studies laboratory at the University of Alberta hospital.

Prior to being discharged patients were prescribed inhaled corticosteroids (ICS), and 1-2mg/kg to 60mg daily prednisone to be used for five days. Following the treatment period patients returned to the ER for a follow-up clinical evaluation. At this time a 'follow-up' urine sample was collected.

To date 38 adult and 31 pediatric pre/post-prednisone urine samples have been collected, catalogued, and analyzed by NMR. An equal number of normal urine samples were randomly selected from the larger normals database (Chapter IV) for comparison to the asthma patients.

VII-B. Sample Preparation

Urine samples were thawed in a biosafety fume hood and a 630 μl aliquot was removed and placed in a 1.5 ml Eppendorff tube followed by the addition of 70 μl of a reference buffer solution ((4.9 mM DSS (disodium-2, 2-dimethyl 2-silapentane-5-sulphonate) and 100 mM imidazole in D_2O) Sigma-Aldrich, Mississauga, ON). Each sample was then brought to a pH of 7.0 +/- 0.1 using HCl and NaOH. An aliquot of 600 μl was taken and transferred to a standard 5 mm glass NMR tube (Wilmad, NJ, USA).

VII-C. NMR Analysis

All ^{1}H-NMR spectra were acquired on a 600 MHz Inova (Varian Inc, Palo Alto, CA.) spectrometer equipped with a 5 mm triple-resonance (HCN) probe with Z-axis gradient coil. One-dimensional ^{1}H-NMR spectra were collected at 25°C with a tnnoesy pulse sequence (one-dimensional, three pulse NOESY, with a transmitter pre-saturation delay of 900 ms for water suppression during the pre-acquisition delay and 100 ms mixing time), and a spectral width of 7200 Hz. The time-domain data points were 64k complex points, acquisition time was 4 s, the 90° pulse was 6.8μs, repetition time was 5 s, four steady state scans, and the number of acquired scans was 32 per FID. The data was apodized with an exponential window function corresponding to a line broadening of 0.5 Hz, zero-filled to 128k complex points, and Fourier transformed [1].

VII-D. Exporting NMR Spectra for Analysis

The exporting of the spectra in XY format, the re-saving of the spectra for Chenomx analysis, and the re-naming of all the files was automated using a macro created by Pascal Mercier, Erik Saude, and Robert Boyko (autoprocess), see Appendix C. The spectra were exported into an ASCII file, which contained two columns: data point in Hz. (X) and the corresponding amplitude (Y). Spectra were exported into 32k data points (32768).

VII-E. Multivariate Statistical Analysis

The xytrace (ASCII format) files of the spectra were sent to Brion Dolenko and Ray Somorjai at the Institute for Biodiagnostics (Winnipeg, MB.) for multivariate statistical analysis.

The prednisone data was categorized into four classes: pediatric pre-prednisone (31), pediatric post-prednisone (31), adult pre-prednisone (38), and adult post-prednisone (38). Classifiers were developed that separated the different classes. For all classifiers, the optimal spectral regions were determined first. This feature selection was carried out by a genetic-algorithm-based feature selection approach[2]. The feature selection was wrapper-based: *i.e.* a classifier, in this case linear discriminate analysis (LDA) with leave-one-out (LOO) internal cross validation. This was used to decide which were the best features.

Having obtained the optimal feature set for each pair classifier, external cross validation (EXV) helped estimate a more realistic error estimate. EXV was carried out by randomly splitting each dataset into a training set and a test set (50:50). The splitting was stratified, *i.e.* the relative proportions of the samples in the two classes were retained in the splits. Details of the statistical classification used can found in a recent textbook[3].

VII-F. NMR Quantification

The methyl singlet of the buffer constituent DSS served as an internal standard for chemical shifts (set to 0 ppm), and for quantification. Spectral identification and quantification was performed using the Chenomx NMR Suite Professional software package Version 3.1 (database available at pH 7.0, Chenomx Inc., Edmonton, AB). Work highlighted in Chapter II has demonstrated that this procedure provides absolute concentration accuracies in excess of 90%[4].

RESULTS

VII-A. Multivariate Classification

Spectra were divided into two portions, aliphatic (0 – 4.7 ppm) and aromatic (5.1 – 11 ppm), with the exclusion of the remaining water peak. The xytrace of the aliphatic and aromatic spectral regions were analyzed for spectral signatures that were able to distinguish the different patient and normal populations (see Methods section above).

Table VII-1 demonstrates the statistical evaluation of the algorithms' ability to distinguish between the different classes using the aliphatic portion of the spectra. For the initial analysis comparisons were drawn between all combinations of normal, adult asthma pre and post-treatment, as well as pediatric asthma pre and post-treatment. Class assignment was defined confident ("crisp") if the sample-assignment probability was greater than 75%. The algorithms were able to distinguish normal subjects from adult asthma pre-treatment (86.2%), adult post-treatment (76.5%), pediatric pre-treatment (77.8%), and pediatric post-treatment (85.7%). Within the disease populations the algorithms were not able to distinguish between the populations very well. Only adult pre-treatment vs. pediatric post-treatment was statistically significant (90.8%). It should be considered that due to the different ages of the populations this might not be a valid comparison.

Classifier Pair	# of Regions	Weight W_1 or W_2	SE Sensitivity	SP Specificity	Overall Accuracy	Crisp SE	Crisp SP	Crisp Accuracy	% Crisp
N vs Adult Pre	7	W2 = 3.5	97.5	93.3	96.6	98.6	96.8	98.3	86.2
N vs Adult Post	6	W2 = 3.25	93.7	92.1	93.4	97.6	95.8	97.3	76.5
N vs Ped Pre	7	W2 = 5.0	96.8	95.6	96.6	98.4	100.0	98.7	77.8
N vs Ped Post	6	W2 = 5.0	97.5	93.5	96.8	98.5	96.2	98.1	85.7
Adult Pre vs Adult post	7	W2 = 1.6	88.9	89.5	89.2	86.2	96.7	91.5	71.1
Adult Pre vs Ped Pre	7	W2 = 1.5	91.1	91.1	91.1	92.3	96.6	94.5	61.1
Adult Pre vs Ped Post	7	W2 = 1.5	95.6	96.8	96.1	97.6	96.4	97.1	90.8
Ped Pre vs Ped Post	7	W2 = 1.0	91.1	93.5	92.1	94.1	90.9	92.9	73.7
Adult Post vs Ped Post	7	W2 = 1.25	94.7	93.5	94.2	100.0	88.2	95.8	69.6
Adult Post vs Ped Pre	7	W2 = 1.65	86.8	91.1	89.2	89.5	96.7	93.9	59.0

Table VII-1. Multivariate diagnosis using aliphatic NMR signatures
a. Multivariate analysis that statistically separates adult pre and post-treatment, as well as pediatric (Ped) pre and post-treatment asthma and normal (N) population using the aliphatic region of NMR spectra.
b. Pre, pre-treatment; Post, post-treatment

Table VII-2 shows the statistical evaluation of the algorithms ability to distinguish between the different classes using the aromatic portion of the spectra (roughly 6.5ppm – 11.0ppm). This included separating normal subjects from adult asthma pre-treatment (89.2%), adult post-treatment (93.4%), pediatric pre-treatment (98.5%), and pediatric post-treatment (97.9%). Within the asthma populations the aromatic region was able to distinguish adult pre-treatment from adult post-treatment (79.5%), adult pre-treatment vs. pediatric pre-treatment (97.8%), adult pre-treatment vs. pediatric pos-treatment (97.4%), pediatric pre-treatment vs. pediatric post-treatment (76.3%), adult post-treatment vs. pediatric post-treatment (100%), and finally adult post-treatment vs. pediatric pre-treatment (94%).

Classifier Pair	# of Regions	Weight W_1 or W_2	SE Sensitivity	SP Specificity	Overall Accuracy	Crisp SE	Crisp SP	Crisp Accuracy	% Crisp
N vs Adult Pre	7	W2 = 2.5	93.7	93.3	93.6	97.2	97.3	97.2	89.2
N vs Adult Post	6	W2 = 1.5	97.5	100.0	98.0	97.4	100.0	97.8	93.4
N vs Ped Pre	4	W2 = 1.35	99.4	97.8	99.0	99.4	97.7	99.0	98.5
N vs Ped Post	3	W2 = 1.35	100.0	96.8	99.5	100.0	96.4	99.5	97.9
Adult Pre vs Adult Post	7	W1 = 2.5	88.9	89.5	89.2	88.2	93.8	90.9	79.5
Adult Pre vs Ped Pre	6	W1 = 1.1	97.8	100.0	98.9	97.7	100.0	98.9	97.8
Adult Pre vs Ped Post	4	W2 = 2.0	100.0	96.8	98.7	100.0	96.6	98.6	97.4
Ped Pre vs Ped Post	6	W2 = 8.0	88.9	83.9	86.8	89.5	85.0	87.9	76.3
Adult Post vs PedPost	3	W1 = 1.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Adult Post vs Ped Pre	3	W1 = 1.0	97.4	97.8	97.6	97.3	97.6	97.4	94.0

Table VII-2. Multivariate diagnosis using aromatic NMR signatures

a. Ability of multivariate analysis to statistically separate adult pre and post-treatment, as well as pediatric (Ped) pre and post-treatment asthma and normal (N) populations using the aromatic region of NMR spectra.
b. Pre, pre-treatment; Post, post-treatment

Spectral regions that allowed for multivariate separation of the disease and normal populations were identified and are presented in Tables VII-3 and 4. Table VII-3 shows the aliphatic regions of data points that contain the unique spectral indices that enable population separation. Often population separation required the identification of seven different regions. Table VII-4 similarly demonstrates the range of data points that contain unique spectral traits that separate the different populations. For analysis of the aromatic regions often less than seven spectral characteristics were required for population separation.

Classifier Pair	1	1	2	2	3	3	4	4	5	5	6	6	7	7
N vs Adult Pre	732	805	2031	2067	2446	2496	2649	2683	2958	3032	3573	3604	3933	3992
N vs Adult Post	585	627	782	805	2035	2051	2475	2535	2708	2732	3739	3791		
N vs Ped Pre	964	1006	1060	1077	1312	1368	2029	2062	2516	2541	3595	3622	3938	3984
N vs Ped Post	798	816	2034	2065	2182	2206	2225	2286	3185	3262	3547	3596		
Adult Pre vs Adult Post	721	764	1144	1165	1685	1721	2352	2382	2386	2430	3466	3490	3766	3807
Adult Pre vs Ped Pre	790	829	1122	1171	1377	1397	2101	2165	2313	2363	2373	2448	2614	2633
Adult Pre vs Ped Post	324	355	1376	1444	1445	1477	2192	2209	3131	3164	3420	3457	3717	3750
Ped Pre vs P Post	391	436	1091	1120	1161	1219	1511	1544	1567	1614	1689	1728	3109	3142
Adult Post vs Ped Post	789	856	1146	1211	1276	1325	1342	1397	1602	1627	1792	1857	3118	3144
Adult Post vs Ped Pre	416	433	788	822	1598	1636	2195	2253	2263	2356	2973	2993	3772	3802

Table VII-3. Aliphatic spectral data points that separate populations

a. Aliphatic data point windows that contain the data to statistically separate adult pre and post-treatment, as well as pediatric (Ped) pre and post-treatment asthma and normal (N) populations using the aromatic region of NMR spectra.

b. Pre, pre-treatment; Post, post-treatment

Classifier Pair	1	1	2	2	3	3	4	4	5	5	6	6	7	7
N vs Adult Pre	878	940	1824	1838	2342	2391	3844	3870	4664	4721	6432	6460	6863	6890
N vs Adult Post	1083	1104	2684	2716	3647	3661	5224	5260	6357	6380	6398	6408		
N vs Ped Pre	1076	1145	3553	3598	5911	5982	6291	6348						
N vs Ped Post	1110	1153	5888	5935	6301	6375								
Adult Pre vs Adult Post	1508	1549	3000	3009	3504	3576	6129	6201	6548	6575	7276	7327	7375	7425
Adult Pre vs Ped Pre	1088	1154	1371	1447	2303	2343	5508	5565	5918	5960	6340	6401		
Adult Pre vs Ped Post	1088	1164	1474	1542	5033	5601	5885	5954						
Ped pre vs Ped Post	1435	1497	2941	2979	4441	4481	4519	4570	6088	6149	7080	7138		
Adult Post vs Ped Post	1092	1154	3076	3144	5921	5951								
Adult Post vs Ped Pre	1088	1161	5879	5982	6365	6439								

Table VII-4. Aromatic spectral data points that separate populations

a. Aromatic data point windows that contain the data to statistically separate adult pre and post-treatment, as well as pediatric (Ped) pre and post-treatment asthma and normal (N) populations using the aromatic region of NMR spectra.

b. Pre, pre-treatment; Post, post-treatment

The data presented in Tables VII-3 and VII-4 are preliminary. Due to the limited sample size it is probable that further analysis, as well as the inclusion of additional samples would reveal shifted, or new regions of interest. Therefore, further NMR interpretation of spectral features and metabolites was not carried out. However, for the purpose of later discussion it is important to note that it was possible to convert data points to actual spectral regions (Hz. or ppm) according to the following equations:

Aliphatic

Eq (1).
$$\begin{aligned} Hz &= 2700.2441426 - 0.2441426 * DataPt. \\ PPM &= 4.5004069 - 0.0004069 * DataPt. \end{aligned}$$

Aromatic

Eq (2).
$$\begin{aligned} Hz &= 5176.5441426 - 0.2441426 * DataPt. \\ PPM &= 8.6276069 - 0.0004069 * DataPt. \end{aligned}$$

DISCUSSION

Metabolomic investigations attempt to identify, quantify, and correlate changes with particular pathophysiologies. Often the disease being studied has numerous pathways in flux and the list of possible compounds to identify can be very large. As well, metabolomics allows for non-hypothesis driven investigations, which may lead to the identification of new metabolites and disease pathways. Multivariate analysis may be the ideal manner in which to perform data reduction to a minimal number of key spectral signatures, which allows for population separation.

Multivariate analysis is capable of analyzing large datasets and reducing information to a level that is easily interpreted by the investigator. Metabolomic investigations often collect data from hundreds of patients or subjects. NMR spectral analysis of the individual subjects contains a large amount of information (32 – 65k). In the study discussed in this appendix the NMR spectra were exported to 32K data points. The data quickly adds up to roughly 1.5 million data points per cohort (pediatric pre-treatment, adult post-treatment, etc.). The identification, quantitation, and statistical analysis of every metabolite, in each urine sample would not be feasible via traditional methods. As well, the pathophysiology of a disease would likely a very complex process requiring the modification of a number of metabolic pathways. Multivariate analysis could be the method by which investigators can handle large datasets, reduce the information to key metabolites (sometimes novel metabolites), and separate distinct populations with statistical significance.

As mentioned in the introduction, there are a few things that must be addressed when performing multivariate analysis. More specific to NMR investigations carried out during the work for this thesis were the issues of constant number of data points, Hz/pt. (hertz per point), DSS reference, H_2O

residual peak, and the final conversion of multivariate signatures to NMR spectra. In Appendix B I will discuss further the identification of novel spectral signatures by advanced NMR techniques and MS.

Data acquisition for NMR involves the digitization of an analogue signal. This digitization process results in data points that are a function of 2^x (where x=1, 2, 3, etc.). Initially the data was collected with 32k (actually 32768, 2^{15}), but now data collection is standardized to 64k data points (65536, 2^{16}). This is important to consider when looking at spectral resolution as well as performing multivariate analysis. During RDP, PCA, or most other methods of analysis the data (multiple spectra) are put into a large matrix and diagonalized to determine the eigenvectors and eigenvalues. It is important that all the spectra being compared have identical number of data points. Related to this is the importance of Hz/pt. In order to allow for similar regions of the spectra to be compared after exporting to the XY format the Hz/pt. must be identical. To achieve this all spectra are acquired with identical sweep width (usually 8012 Hz) and number of data points (64k) while acquisition time is then set by the instrument (roughly 4 s).

It is important to ensure that the reference standard (DSS for my studies) was properly referenced before spectra were exported. Following referencing the spectrum it is important to remember that although the NMR chemical shift scale of ppm has been referenced to 0, the data points will extend past the reference peak to the outer portions of the spectrum. Furthermore, it should be understood that because the reference peak is now situated at 0 ppm, the peak is data point 56203 along the 64k data conversion scale. Other peaks common to NMR are important to note when performing multivariate analysis. The residual water peak, roughly 4.7 ppm, will vary slightly from spectrum to spectrum. Although this is largely not due to the physiology of the subject and the sample multivariate analysis will identify this peak as being the primary source of difference between spectra. This is something that can be easily corrected by removing the water peak from the data set (xy trace) before performing the statistical analysis. Also, as mentioned in Chapter IV, subject hydration will alter urine metabolite concentrations. It is important that the statistical tool identify the creatinine peaks (data points 41236 – 41481, and 36246 – 36574) and standardize each individual spectrum before performing the multivariate analysis.

This chapter provides a brief outline of a continuing study that attempts to initiate the use of NMR to analyze human asthma urine to diagnose and track asthma improvement and treatment through the identification and quantification of key metabolites. Patient recruitment, compound identification, and multivariate analysis are still proceeding at the time this thesis was written so a final conclusion of the study was not possible. However, the study highlights some important findings that should be considered.

The primary advantage of multivariate statistical analysis is the identification of unique and novel spectral signatures that may lead toward the discovery of new metabolites and relevant metabolic pathways. This could prove to be useful for basic research as it expands the current understanding of disease pathophysiology. In addition, drug development could benefit greatly since the physiological impact of new pharmaceuticals is often not known. Metabolomic investigations could allow for the identification of specific pharmaceutical metabolites and the metabolic pathways impacted by the new therapies.

Although multivariate analysis often requires collaborations with experts in the statistical or engineering fields to ensure it is performed correctly, the possibilities of the analysis are encouraging. Work is ongoing to optimize the capabilities and ability to identify new metabolites that could prove useful in the understanding of disease pathophysiology, and thus further the development of improved patient therapeutics.

LITERATURE CITED

1. Kumar, A., R.R. Ernst, and K. Wuthrich, A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem Biophys Res Commun, 1980. 95(1); 1-6.
2. Nikulin, A.E., et al., Near-optimal region selection for feature space reduction: novel preprocessing methods for classifying MR spectra. NMR Biomed, 1998. 11(4-5); 209-16.
3. Somorjai, R.L., et al., A Data-Driven, Flexible Machine Learning Strategy for the Classification of Biomedical Data. Artificial Intelligence Methods and Tools for Systems Biology, Computational Biology Series, ed. W. Dubitzky and F. Azuaje. Vol. 5. 2004: Springer. 67-85.
4. Saude, E.J., C.M. Slupsky, and B.D. Sykes, Optimization of NMR analysis of biological fluids for quantitative accuracy. Metabolomics, 2006. 2(3); 113-123.

CHAPTER VIII

Pediatric asthma longitudinal study

OVERVIEW

Chapter VII provided a brief description of a continuing study of human asthma. The initial project chose acute and treated asthma because it was believed that the two extremes were physiologically distinct enough to provide the best separation clinically and differential metabolic profiles. Perhaps a more difficult consideration is not whether a diseased and a control population can be distinguished, but rather can a disease process be followed through all the subtleties of the clinical presentaion. *This chapter briefly describes a research project that worked to identify and distinguish slight metabolic alterations following the pharmacological maintenance of asthma.* Pediatric asthma patients who frequented the outpatient clinic at the UofA hospital were sampled over subsequent visits. These patients were also followed through changes to their medication. Following a single individual over a period of two years could produce metabolic indicators that correlate with the proper treatment of asthma, or metabolites that are predictive of future exacerbations.

INTRODUCTION

Investigations toward a fuller understanding of particular disease processes often involve the comparison of normal and diseased groups. Chapters

IV - VII highlighted studies that provided a metabolic definition of a normal human and guinea pig population, as well as examples of comparative metabolomic studies for asthma and CF. As the studies progressed the guiding question changed from 'What are the gross metabolic and associated physiological differences between the two populations?' to perhaps a more clinically relevant question 'What are the different stages or disease processes underlying the broader patient diagnosis?'. This chapter highlights on-going work attempting to identify metabolic and immunological pathways responsible for, and influenced by, the progression and treatment of asthma.

Chapter VII demonstrated the initial investigation into human asthma metabolomics through the sampling of urine. The preliminary study focused on two distinct populations; acute emergency asthma patients and control subjects (for one study the controls were patients subjected to two-weeks of steroid treatment) It was hypothesized that by choosing these distinct populations differences in metabolic and immunological pathways would be easier to identify. This identification would provide insight into the pathways affected by, and those producing, the acute asthma episode. It was readily acknowledged that in order to determine a better understanding of the basic mechanisms underlying asthma it would be useful to identify the subtle changes in stable asthma patients.

Asthma is a disorder that has episodes of severe acute exacerbation, as well as a sustained level of inflammation that can cause significant physiological damage and pulmonary remodeling to the lungs over time[1-5]. Therapeutic intervention attempts to limit the number and severity of acute episodes for patients and tries to manage the continual state of inflammation[1, 6]. The pediatric asthma and allergy outpatient clinics at the Universities of Alberta and Manitoba care for sick children at varying stages of treatment for pulmonary disorders. Pulmonary function tests and allergen skin prick testing were conducted in addition to standard medical care in order to assess treatment efficacy. Pediatric patients provided urine samples that were analyzed by 1D ^{1}H-NMR. NMR was used to identify metabolic pathways unique to asthma patients, as well as the metabolites that might respond to therapeutic measures in order to obtain a more detailed understanding of asthma [7]. The metabolic understanding of baseline asthma, acute exacerbation, and the indicators that precede the acute exacerbation is necessary as this metabolic information provides an adjunct to medical decisions regarding appropriate therapeutic intervention[8].

EXPERIMENTAL PROCEDURES

VIII-A. Pediatric Patient Recruitment

Informed consent was obtained from asthma patients in accordance with guidelines established by the University of Alberta Health Research Ethics Board. Presently 61 initial pediatric samples, with a total of 130 pediatric samples have been collected as of July 2006 (multiple samples per patient). Some patients have had as many as six urine samples collected during repeat visits to the UofA pediatric outpatient clinic. All samples were immediately stored at -20°C at the clinic and then brought to the laboratory where sodium azide was added to a final concentration of 2.5mM and stored in a deep-freeze (-80°C).

In addition, samples collected at the University of Manitoba Childrens Hospital included an additional 202 pediatric asthma patients, as well as 267 pediatric control urine samples. All samples were stored at -20°C at the clinic and then placed in a deep-freeze (-80°C) until they were shipped to the NANUC. The samples were shipped with dry ice and remained frozen during transport. The samples were immediately transferred to the NANUC deep-freeze.

For a preliminary analysis, 38 children aged 4 to 16 were selected from the larger pediatric outpatient database. Two urine samples per patient, the initial and follow-up visit, were selected and analyzed (see below). In addition, an equal number of control pediatric samples were randomly selected and analyzed. For the preliminary analysis 38 pediatric asthma patients were separated into two groups. Patients who showed improvement based upon clinical history, exam, and pulmonary function tests between the first and second visits were identified as better. Those that showed a decline in the management of asthma were termed worse.

VIII-B. Sample Preparation

Urine samples were thawed in a biosafety fume hood. A 630 µl aliquot was taken and placed in a 1.5ml Eppendorff tube followed by the addition of 70 µl of a standard solution (4.9mM DSS {disodium-2, 2-dimethyl 2-silapentane-5-sulphonate} Sigma-Aldrich, Mississauga, ON). Each sample was brought to a pH of 7.0 +/- 0.1. An aliquot of 600µl was taken and transferred to a standard 5mm glass NMR tube (Wilmad, NJ).

VIII-C. NMR Analysis

All ^{1}H NMR spectra were acquired on a 600 MHz Inova (Varian Inc, Palo Alto, CA.) spectrometer equipped with a 5mm triple resonance probe with z-axis pulsed field gradients. One-dimensional ^{1}H NMR spectra were collected at 25°C with a tnnoesy pulse sequence (one-dimensional NOESY with transmitter saturation during the preacquisition delay and 100ms mixing time) with a spectral width of 8200 Hz, four steady state scans preceding acquisition, and 32 transients were acquired for each spectrum which was apodized with an exponential decay corresponding to a line broadening of 0.5 Hz prior to Fourier Transformation[9]. Spectra were zero-filled to 128k. The acquisition time per scan was 4s, with a preacquisition 900ms for a total recycling time of 5s. Presaturation had an effective γB_1 field of 60Hz.

VIII-D. Multivariate Analysis

NMR spectra were exported to XY format using the macro autoprocess as described in Appendix A. Each spectrum was decomposed into 64k data points (65536) in text format. The spectral xytrace files were sent to Brion Dolenko and Ray Somorjai at the Institute for Biodiagnostics (Winnipeg, MB.) for feature selection, which was wrapper-based: *e.g.* a classifier, in this case linear discriminate analysis (LDA) with leave-one-out (LOO) internal cross validation, was used to decide which features were the best.

Once the optimal feature set for each pair classifier was obtained, external cross validation (EXV) helped estimate a more realistic error estimate. EXV was performed by splitting each dataset, randomly, into a training set and a test set (50:50). The splitting was stratified, *i.e.* the relative proportions of the samples in the two classes were retained in the splits. Details of the statistical classification used can be found in a recent textbook[10].

VIII-E. NMR Quantification

The buffer constituent DSS served as internal standard for spectral quantification, while the methyl resonance peak of DSS was utilized as an internal reference set to 0 ppm. Spectral quantification of chemical constituents utilized Chenomx NMR Suite Professional software package Version 3.1 (database available at pH 7.0, Chenomx Inc., Edmonton, AB). Urine metabolites were normalized to internal creatinine concentrations in order to correct for dilution or hydration effects.

VIII-F. Statistical Analysis

Statistical significance of the urine metabolites was determined by performing an analysis of variation (ANOVA). A p-value < 0.05 was determined to be statistically significant.

RESULTS

VIII-A. Pulmonary Function

As part of the clinical assessment, pediatric patients produced a FEV1 (forced expiratory volume in 1s) for each visit. Improvements in FEV1 correlated with the patient feeling better on the second visit. The inverse was also true, those patients that felt worse on their second visit showed a significant reduction in FEV1. Figure VIII-1 demonstrates the correlated rise and fall of FEV1 with a patients' improvement or deterioration upon the second visit, respectively.

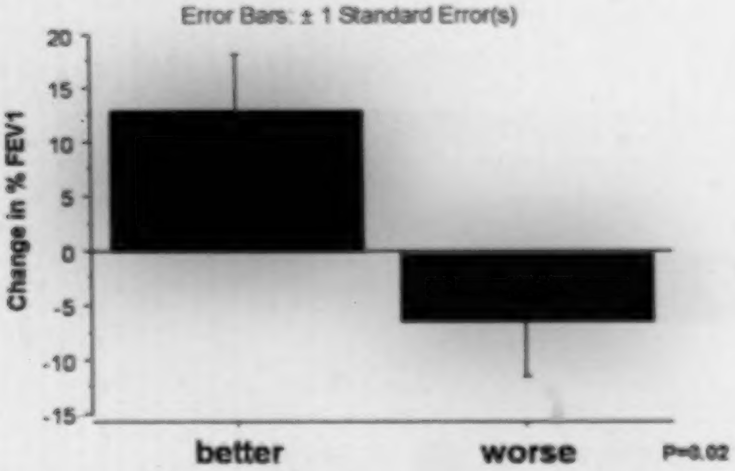


Figure VIII-1. **FEV1 correlates with clinical presentation between visits.** FEV1 analysis was performed on the first and second visit, which correlated with the patients' improvement or deterioration.

VIII-B. NMR Metabolite Quantitation

Preliminary spectral analysis by a summer student Chris Skappak included 38 easily identifiable metabolites. The only metabolite to reach statistical significance, between the better and worse patients, was 1-methylhistamine.

Normalized 1-methylhistamine showed a significant reduction in concentration in the second sample collected from patients who presented to the clinic as better. Conversely, patients whose lung function deteriorated leading up to the second visit demonstrated elevated levels of 1-methylhistamine in the urine. Figure VIII-2 demonstrates the difference in normalized 1-methylhistamine from the follow-up and the initial visit (day two minus day one).

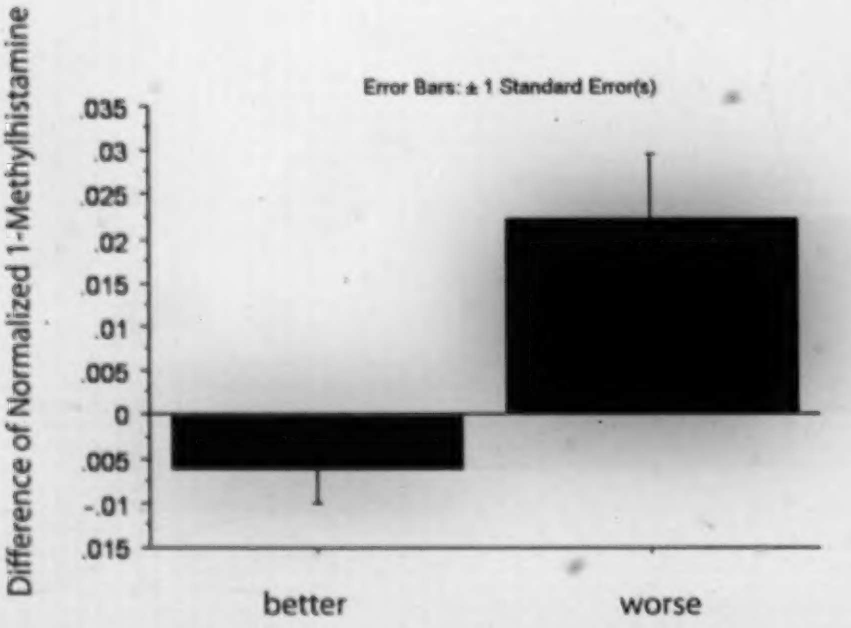


Figure VIII-2. **1-methylhistamine correlates with clinical follow-up**. This figure shows the difference in the normalized 1-methylhistamine levels in the follow-up compared to the initial urine sample for patients who were better or worse during the second visit.

The level of 1-methylhistamine was elevated in the initial visit urine sample from pediatric patients who presented as worse on the follow-up visit when compared to patients that presented as better on the follow-up visit. Figure VIII-3 demonstrates the elevated level of normalized 1-methylhistamine on the initial visit in patients who presented as worse at the follow-up visit to the clinic.

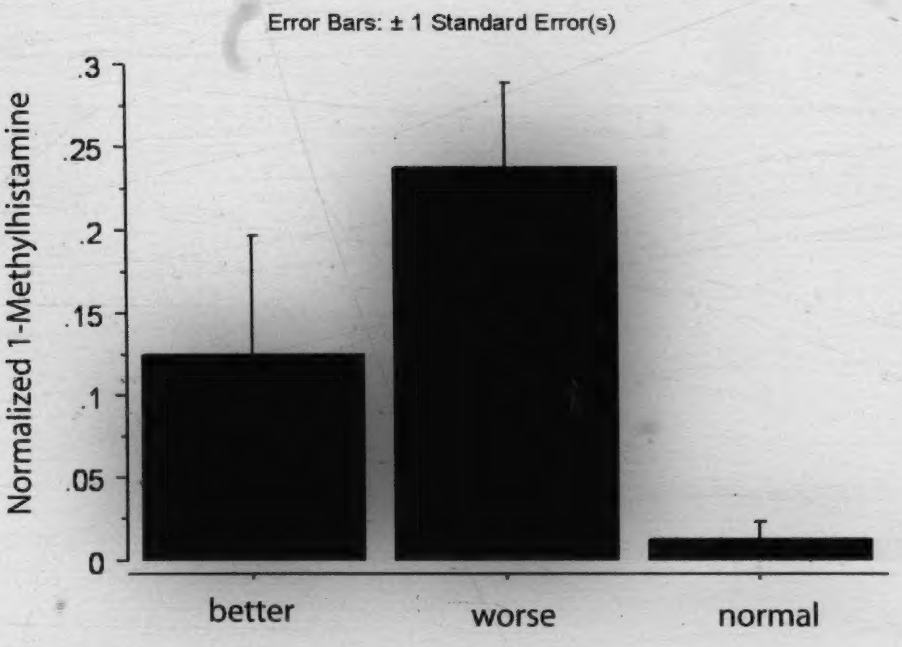


Figure VIII-3. **1-methylhistamine predicts follow-up presentation**. Normalized 1-methylhistamine was significantly increased in the initial urine sample for patients that presented as worse at the follow-up evaluation.

VIII-C. Multivariate Separation

Although multivariate analysis was able to distinguish pediatric asthma patients between their initial and subsequent visits, the final version of the multivariate analysis was not returned in time for documentation in this thesis.

DISCUSSION

Although this study is still in the preliminary phase a significant amount of pediatric asthma urine samples were collected and additional samples will continue to be analyzed by NMR and multivariate statistical analysis. This initial investigation demonstrated the ability of NMR to identify the immunologically relevant metabolite 1-methylhistamine. With continued analysis, and the return of the multivariate analysis, additional metabolites might be identified that could serve as biomarkers for pediatric asthma patients who present to the clinic with worse pulmonary function.

Higher normalized urinary levels of 1-methylhistamine on the follow-up visit, when compared with the initial visit, correlated with a patients' worse presentation (Figure VIII-2). A worse follow-up presentation also correlated significantly with a reduction in pulmonary function (FEV1, Figure VIII-1). It appears that 1-methylhistamine served as a biomarker for the reduction in pulmonary function. Also of interest is that a greater normalized level of 1-methylhistamine in the initial visit urine sample correlated significantly with a patients' likelihood of presenting as worse on the follow-up visit (Figure VIII-3). This suggests that 1-methylhistamine could be prognostic and could warn of a patients' deteriorating pulmonary function. This work is being confirmed by MS as well.

The cause for elevated 1-methylhistamine in worsening asthmatic lungs could be because only 2 – 3% of histamine is excreted in the urine as unchanged histamine. Seventy percent is converted to 1-methylhistamine (10% of which is excreted in the urine), while the rest is converted to methylimidazole acetic acid[11-13]. Histamine is known to be a mediator of acute allergic reactions, and was one of the first chemical substances to be associated with mast cells [14-20]. Histamine is a potent bronchospastic and a vasodilator and is known to be associated with allergic reactions and asthma[21-24].

This study will continue with the multivariate analysis and the identification of additional metabolites. The metabolites identified by the multivariate analysis should provide both a diagnostic use, as well as provide a new understanding of the pathways responsible for the unique physiology found in asthma. With a large longitudinal database complete with pulmonary function tests, prescribed treatments, as well as sequential urine samples from patients this project should enjoy continued success.

LITERATURE CITED

1. Canonica, G.W., Treating asthma as an inflammatory disease. Chest, 2006. 130(1 Suppl); 21S-28S.
2. Eum, S.Y., et al., Inhibition of allergic airways inflammation and airway hyperresponsiveness in mice by dexamethasone: role of eosinophils, IL-5, eotaxin, and IL-13. J Allergy Clin Immunol, 2003. 111(5); 1049-61.
3. Hogan, S.P., et al., Aeroallergen-induced eosinophilic inflammation, lung damage, and airways hyperreactivity in mice can occur independently of IL-4 and allergen-specific immunoglobulins. J Clin Invest, 1997. 99(6); 1329-39.
4. Kim, J., et al., Prevention and reversal of pulmonary inflammation and airway hyperresponsiveness by dexamethasone treatment in a murine model of asthma induced by house dust. Am J Physiol Lung Cell Mol Physiol, 2004. 287(3); L503-9.
5. Laitinen, L.A., et al., Eosinophilic airway inflammation during exacerbation of asthma and its treatment with inhaled corticosteroid. Am Rev Respir Dis, 1991. 143(2); 423-7.
6. Irwin, R.S. and N.D. Richardson, Side effects with inhaled corticosteroids: the physician's perception. Chest, 2006. 130(1 Suppl); 41S-53S.
7. Cicutto, L.C. and G.P. Downey, Biological markers in diagnosing, monitoring, and treating asthma: a focus on noninvasive measurements. AACN Clin Issues, 2004. 15(1); 97-111.
8. Green, R.H., et al., Asthma exacerbations and sputum eosinophil counts: a randomised controlled trial. Lancet, 2002. 360(9347); 1715-21.
9. Kumar, A., R.R. Ernst, and K. Wuthrich, A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem Biophys Res Commun, 1980. 95(1); 1-6.

10. Somorjai, R.L., et al., A Data-Driven, Flexible Machine Learning Strategy for the Classification of Biomedical Data. Artificial Intelligence Methods and Tools for Systems Biology, Computational Biology Series, ed. W. Dubitzky and F. Azuaje. Vol. 5. 2004: Springer. 67-85.

11. Uvnas, B., C.H. Aborg, and A. Bergendorff, Storage of histamine in mast cells. Evidence for an ionic binding of histamine to protein carboxyls in the granule heparin-protein complex. Acta Physiol Scand Suppl, 1970. 336; 1-26.

12. Kapeller-Adler, R., Histamine catabolism in vitro and in vivo. Fed Proc, 1965. 24; 757-65.

13. Snyder, S.H. and J. Axelrod, Tissue metabolism of histamine -C14 in vivo. Fed Proc, 1965. 24; 774-6.

14. Schulman, E.S., et al., Heterogeneity of human mast cells. J Immunol, 1983. 131(4); 1936-41.

15. Windaus, A. and W. Vogt, Synthese des imidazolyathylamins. Ber Dtsch Chem Ges, 1907. 3; 3691-3695.

16. Fox, C.C., et al., Isolation and characterization of human intestinal mucosal mast cells. J Immunol, 1985. 135(1); 483-91.

17. Benyon, R.C., M.A. Lowman, and M.K. Church, Human skin mast cells: their dispersion, purification, and secretory characterization. J Immunol, 1987. 138(3); 861-7.

18. Schwartz, L.B. and K.F. Austen, Structure and function of the chemical mediators of mast cells. Prog Allergy, 1984. 34; 271-321.

19. Church, M.K., et al., Studies into the mechanisms of dermal inflammation using cutaneous microdialysis. Int Arch Allergy Immunol, 1997. 113(1-3); 131-3.

20. Dreskin, S.C., M.A. Kaliner, and J.I. Gallin, Elevated urinary histamine in the hyperimmunoglobulin E and recurrent infection (Job's) syndrome: association with eczematoid dermatitis and not with infection. J Allergy Clin Immunol, 1987. 79(3); 515-22.

21. Dale, H.H. and P.P. Laidlaw, The physiological action of beta-iminazolylethylamine. J Physiol, 1910. 41(5); 318-344.

22. Best, C.H., et al., The nature of the vaso-dilator constituents of certain tissue extracts. J Physiol, 1927. 62(4); 397-417.

23. Lewis, T., The blood vessels of human skin and their responses. 1927, London: Shaw & Son.

24. Riley, J.F. and G.B. West, Skin histamine; its location in the tissue mast cells. AMA Arch Derm, 1956. 74(5); 471-8.

CHAPTER IX

Conclusion

OVERVIEW

The final chapter of this thesis provides a unifying discussion regarding the application of NMR spectroscopy toward metabolomic and clinical research. *This final discussion begins with an overview of the previous eight chapters, the major conclusions for each, and their contributions to the growing field of metabolic research.* The chapter also provides additional comments regarding some of the directions the field of metabolomics and clinical research may be heading.

SUMMARY

This thesis has presented research that is not typical to the historical application of NMR. The broader field of basic metabolic research has been a focus of scientific research for hundreds of years[1]. More recently the field of metabolomics has created a resurgence in the interest of metabolism and the wealth of information that may be garnered from the study of metabolic pathways[2]. The relationship between metabolism, phenotype, and pathophysiology has initiated a philosophical shift from reductionism to a more global analysis that has captured the interest of many clinical researchers[3-8]. The possibility of detecting and tracing key metabolites is of primary interest to clinical researchers because of the close connection between active metabolic pathways and the continually changing physiology of a diseased individual[9, 10].

Although a powerful tool for protein structure determination, NMR has many features that distinguish itself as an application suited for metabolic research. The basic theories and technology have been around for decades and advancements in the field continue to increase the breadth of NMR applications. Today, there is a wealth of acquisition experiments available to NMR; however, 1D ^{1}H-NMR remains the workhorse of spectroscopy[11-14]. Against a backdrop of different nuclei, mutli-dimensional pulse sequences, solid state NMR, and high-powered pulses, the 1D proton spectrum is deceivingly simple. Chapter II documents research conducted that optimized the efficacy of 1D proton spectrum acquisition and quantitative analysis[15]. Appropriate acquisition parameters and quantitative software have significant effects on metabolite concentration accuracy. As well, engineered hardware and quantum physical phenomena, such as spin-lattice (T_1) relaxation, had to be considered to ensure final accuracy during analysis.

A biochemical continuation of this work is shown in Chapter III where the importance of sample preparation and storage procedures to the preservation of metabolites in the urine was demonstrated. This research was of particular importance to human studies where urine samples are collected at separate locations, often as an additional step during a medical exam or treatment.

Following collection, the urine samples are stored briefly on site, transported, and stored again at the NMR lab for a period ranging from a couple weeks, for data acquisition, and up to years, as part of a continuing biological database. Due to the nature of urine and the inevitability of bacterial contamination Chapter III demonstrates the appropriate measures that must be taken to ensure metabolite fidelity prior to data acquisition.

An intriguing addition to my research, which turned out to be a major focus of my thesis, was the normal study documented in Chapter IV. Scientific analysis involves the measurement of a treated phenomena followed by comparison with a baseline or control measurement. In human and clinical research the selection of control subjects is often based upon proximity, if the subject is in the lab, is a friend, or a relative they are sampled (while attempting to limit ethical conflicts). Chapter IV catalogues the large degree of metabolite variability in the urine and initiates some thought regarding biological normality, homeostasis, and appropriate statistical tools. Chapter V documents my first application of NMR to clinical and metabolomic research. NMR analysis of sputum samples from CF patients revealed indices of immunological effector cell infiltration and activation[16]. Although this was the first study carried out in the field of metabolomics by myself and collaborating laboratories, it opened the door to the possibilities of metabolic research in biological samples to understand unique disease processes. To understand pulmonary pathophysiology further a study was conducted by sampling urine from a guinea pig model of asthma, as described in Chapter VI. Unlike human studies, the animal model allows for a controlled environment and disease state with minimal environmental influence and an absence of pharmaceuticals. The guinea pig study provided detailed pulmonary function information and cell count differentials, as well as an opportunity to develop a multivariate statistical analysis tool for population separation based upon urine spectra. Work continues to detail the metabolic pathways indicative of the different pulmonary states associated with allergy and asthma.

The transition to human disease metabolomics as sampled through the urine began with a study of acute asthma patients prior to and following treatment with corticosteroid (Chapter VII). The sampling of acute patients admitted to the emergency department allowed for a study of 'active' asthma patients. Although a complete biochemical analysis was not completed by the time of this thesis, multivariate statistical analysis was able to separate the different patients based upon NMR spectral signatures from their respective urine samples. This study was the first to use NMR as an analytical tool to investigate urine samples and distinguish asthma patients prior to and following treatment, as well as from control samples. Future work will provide a biochemical understanding of the unique spectral signatures indicative of the acute and treated subjects.

Although it is a unique technique to diagnose acute asthma patients by NMR analysis of the urine it was thought a greater challenge to follow the more subtle metabolic changes of stable asthma patients. Chapter VIII documents a longitudinal study where pediatric outpatients provided samples during follow-up visits. The metabolomic investigation allowed for the identification of compounds that changed during episodes of severe asthma and different therapeutic medications. It is believed that with continuing work these metabolic changes could provide indices of therapeutic efficacy as well as prognostic markers of imminent acute asthmatic episodes.

DISCUSSION

The research for this thesis commenced roughly around the same time as the first release of the Human Genome Project in 2001[17]. The initial excitement about the work accomplished by the project and the newfound knowledge was felt throughout the scientific community. There was also a growing excitement toward the possibilities for future studies. That same anticipation launched similar initiatives in proteomics and helped push the creation of the terms metabonomics and metabolomics. My interest in human and clinical research drew me to a metabolic focus. In addition, my instruction in NMR provided me with an ideal tool through which to perform metabolomic investigations. This thesis follows my development of NMR spectroscopy as a tool for metabolomic studies, as well as my attempts to study normal metabolism and disease processes in pulmonary diseases.

NMR has become a primary tool for metabolomic investigations. Chapter I highlighted aspects that should be optimized to ensure accurate data collection and quantitative analysis. As the technology improves standard tests should continue to ensure that the accuracy remains above 90%. For metabolomic studies new techniques, such as in-line LC-NMR, have created new possibilities for the detection and resolution of metabolites by separating compounds before they enter the spectrometer. For these experiments solvent suppression becomes a priority; to solve this some pulse sequences, such as WET, can be modified to remove different types of solvents[18]. Robotics and automated sample changers have also been introduced into the field to assist with the large number of samples required for metabolomic studies. However, the area of robotics is progressing slowly due to issues such as programming and the reproducibility of data acquisition. NMR continues to be one of the leading technologies for metabolomic research, and further developments in hardware and software will continue to enhance the applicability of NMR for metabolic studies.

The qualitative and quantitative ability of NMR spectroscopy lead to the discovery that urine metabolites change within the same sample over time. Similar to NMR optimization, sample preparation is a laboratory step that is often overlooked. However, it is important to recognize that it has serious implications for final analysis. Chapter II describes a study into the degree of metabolite change and the effects of preparation and storage on the fidelity of urine samples. Certainly the subject providing the sample, the scenario surrounding collection, and the sample itself has implications on the degree of change during the analytical process. It was important to my research that I establish the proper procedures for clinical studies with urine. Investigators should also consider the use of standardized procedures in similar studies, as well; these techniques should be implemented with every project that includes a new type of sample. Often metabolomic studies involve additional chemical or physical preparatory steps. Some studies perform LC-MS, while other studies sample whole tissues followed by grinding and extraction (acetonitrile or perchloric acid extraction). Each of these rather intensive laboratory procedures can have important effects on the metabolites identified during final analysis. My sample preparation study (Chapter III) demonstrated the significant changes that can take place due to bacterial contamination, or simply through non-enzymatic degradation; samples that require extensive physical and chemical laboratory preparation can influence the metabolic profile even greater. The Metabolomics Society has begun to standardize both laboratory and reporting methods for metabolic studies, which will offer direction to future research practices.

One area that is beginning to attract attention is MRS, which eliminates the need to sample a research subject [19, 20]. For human studies a MRI machine is used to detect specific molecules that are important for a particular disease process. Traditionally MRI is used in the clinical setting to identify larger physiological changes associated with tissues and organs. MRS focuses the analysis to key metabolites and biomarkers in the human body that are correlated with a disease. MRS allows for metabolite identification and the tracking of particular metabolic pathways within the human body. Although the technique is still being developed, future work could allow for a large number of metabolites to be identified within the human body, thereby circumventing many of the issues that arise as a result of biological sampling[21].

One important issue I dealt with was deciding upon the appropriate sample type. I performed NMR analysis on ground tissue, ground tissue following perchloric acid extraction, cardiomyocytes, cell cultures, cell culture supernatant, blood, serum, urine, sputum, bronchoalveolar lavage, and breath condensate. Chapter V outlines my investigation of CF through the use of sputum samples. Although sampled directly from the lung, sputum is difficult to collect, and a 'plug' must be separated from the liquid phase of the sputum. In addition, due to the thick nature of sputum (especially for CF patients) the NMR spectra were difficult to analyze. On the other hand urine collection for other studies such as pneumonia and asthma allowed for easy collection of a biofluid from patients. In addition, sputum is nearly impossible to obtain from some of the patients; *e.g.* the elderly pneumonia patients and pediatric asthma patients. The quintessential dilemma for my thesis research was between the direct sampling nature of sputum, though difficult to obtain, and the ease of urine collection, which is a 'distant' biofluid removed from the pulmonary space. Some questions remain regarding the transition of pulmonary metabolites to the urine (i.e. dilution and modification), as well as the feasibility of 'intermediate fluids' such as blood. A study was drafted to collect human sputum, blood, and urine (I ran a similar investigation as part of a mouse pneumonia study with Dr. Lacy) to follow metabolites from the lung, through the blood, and into the urine. Unfortunately, time did not allow for this human study to be completed; however, this would be helpful to find an appropriate balance between ease of sample collection and sampling invasiveness for pulmonary metabolomic investigations.

Using urine as the primary biofluid for the majority of my PhD. research I was surprised to discover how little is known about urine. On a purely clinical basis a lot of research has focused on the implications of ion transport and general filtration (creatine and creatinine) from a physiological standpoint, but very little was known about urine metabolites. With the support of Dr. Marrie I established a normals study, which revealed a significantly large degree of variability of urine metabolites in normal urine. To satisfy the variability of the larger population it was believed that each individual must maintain a restricted metabolite window within the larger variation achieved by the population. In fact almost every subject was able to traverse the population metabolite distribution. This variation was generally maintained for all metabolites detected, including amino acids, citric acid cycle intermediates, and other dietary metabolites. If the variability was not the result of poorly chosen metabolites the next philosophical question was "How does one reconcile between homeostasis and such a large normal variation?". The additional data from the control guinea pigs, with many of the biological, environmental, and dietary factors controlled, produced similarly large fluctuations in metabolite variation. I believe this variability is true, is not an aberration, and might be a reflection of the biological nature of urine. Urine is collected filtration of the blood (representing metabolites from all over the body), and is stored in the bladder for hours. Therefore urine could represent biological processes that occurred over the period of a few hours throughout the body. For comparison, other metabolites are known to vary to a lesser degree in other biofluids, such as blood (or over 24 hr. analysis). The ability to control blood metabolite concentrations is maintained by removing excess to the surrounding cells or through the kidneys. However, having shown the large degree of metabolite variability it was intriguing that there was an identical population mean for two metabolites in a normal study carried out in Italy[22]. This raises the possibility that what my study observed was, in fact, 'normal'. Even though the degrees of variability are large, perhaps diseased processes will still overshadow such large variation.

Basic researchers and clinicians have observed correlations between biological processes and the pathophysiology of a disease. Therefore, even with the large human metabolic variability, it is possible to identify disease processes and make general statements regarding perturbed physiology (Chapters VII and VIII). However, in human metabolomic investigations, how do we identify the underlying disease processes amongst the highly variable normal human physiology?

The degree of metabolite variability found in the normal study raised some interesting questions regarding appropriate methods of statistical analysis. I recently began working with Dr. Brian Rowe, as well as a clinical statistician, Dr. Akhmetshin, to review current methods of statistical analysis and how human variation and metabolic behaviour should be expressed. Appendix A briefly addresses some of the issues pertaining to statistical analysis. The observed metabolite variation does not conform to basic statistical assumptions found in statistical tools often used by researchers (*e.g.* non-parametric vs. parametric analysis, Anova and student t-test). Another philosophical and statistical question remains "How best do we represent human variability, and within such variability, how do we properly identify significant differences?".

Metabolomics represents an attempt to capture all the metabolic information possible from a biological system by sampling and identifying as many unique molecules and active pathways as possible. The raw data can then be quantified, statistically interpreted, and relevant pathways described in an appropriate manner to provide unique information about a biological stress or disease state. The multi-disciplinary requirements of metabolomic investigations offers tremendous opportunities for collaboration and the further development of scientific skills. I have tried to assist in the development of the tools required for metabolomic research, as well as further the understanding of asthma and other diseases. The work presented in this thesis has demonstrated some of the appropriate procedures for metabolomic investigations, and hopefully has stimulated philosophical and scientific discussions surrounding these studies. Through careful thought and the implementation of rigorous analytical techniques, metabolomic investigations will continue to help elucidate disease states through the understanding of metabolic pathways.

LITERATURE CITED

1. German, J.B., B.D. Hammock, and S.M. Watkins, Metabolomics: building on a century of biochemistry to guide human health. Metabolomics, 2005. 1(1); 3-9.
2. Watson, A.D., Thematic review series: Systems Biology Approaches to Metabolic and Cardiovascular Disorders. Lipidomics: a global approach to lipid analysis in biological systems. J Lipid Res, 2006. 47(10); 2101-11.
3. Joyce, A.R. and B.O. Palsson, The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol, 2006. 7(3); 198-210.
4. Mayeno, A.N., R.S. Yang, and B. Reisfeld, Biochemical reaction network modeling: predicting metabolism of organic chemical mixtures. Environ Sci Technol, 2005. 39(14); 5363-71.
5. Nicholson, J.K., et al., Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov, 2002. 1(2); 153-61.
6. Porterfield, D.M., Measuring metabolism and biophysical flux in the tissue, cellular and sub-cellular domains: Recent developments in self-referencing amperometry for physiological sensing. Biosens Bioelectron, 2006.
7. Sabatine, M.S., et al., Metabolomic identification of novel biomarkers of myocardial ischemia. Circulation, 2005. 112(25); 3868-75.
8. Tsang, T.M., et al., Metabolic profiling of plasma from discordant schizophrenia twins: correlation between lipid signals and global functioning in female schizophrenia patients. J Proteome Res, 2006. 5(4); 756-60.
9. Brindle, J.T., et al., Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med, 2002. 8(12); 1439-44.
10. Coen, M., et al., Proton nuclear magnetic resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis. Clin Infect Dis, 2005. 41(11); 1582-90.

11. Hore, P.J., Solvent suppression in fourier transform nuclear magnetic resonance. Journal of Magnetic Resonance, 1983. 55; 283 - 300.
12. Lenz, E.M., et al., A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. J Pharm Biomed Anal, 2003. 33(5); 1103-15.
13. Potts, B.C., et al., NMR of biofluids and pattern recognition: assessing the impact of NMR parameters on the principal component analysis of urine from rat and mouse. J Pharm Biomed Anal, 2001. 26(3); 463-76.
14. Van, Q.N., G.N. Chmurny, and T.D. Veenstra, The depletion of protein signals in metabonomics analysis with the WET-CPMG pulse sequence. Biochem Biophys Res Commun, 2003. 301(4); 952-9.
15. Saude, E., C. Slupsky, and B. Sykes, Optimization of NMR analysis of biological fluids for quantitative analysis. Metabolomics, 2006. 2; 113-23.
16. Saude, E.J., et al., NMR analysis of neutrophil activation in sputum samples from patients with cystic fibrosis. Magn Reson Med, 2004. 52(4); 807-14.
17. Lander, E.S., et al., Initial sequencing and analysis of the human genome. Nature, 2001. 409(6822); 860-921.
18. Smallcombe, S.H., Patt, S. L., and Keifer, P. A., WET Solvent Suppression and Its Applications to LC NMR and High-Resolution NMR Spectroscopy. Journal of Magnetic Resonance, 1995. Series A 117; 295 - 303.
19. Barrett, T., et al., Macromolecular MRI contrast agents for imaging tumor angiogenesis. Eur J Radiol, 2006. 60(3);353-66.
20. Chernov, M.F., et al., Proton magnetic resonance spectroscopy (MRS) of metastatic brain tumors: variations of metabolic profile. Int J Clin Oncol, 2006. 11(5); 375-384.
21. Woods, J.C., et al., Hyperpolarized (3)He diffusion MRI and histology in pulmonary emphysema. Magn Reson Med, 2006. 56(6); 1293-300.
22. Zuppi, C., et al., 1H NMR spectra of normal urines: reference ranges of the major metabolites. Clin Chim Acta, 1997. 265(1); 85-97.

APPENDIX A

Multivariate analysis of NMR metabolomic data

OVERVIEW

Multivariate statistical analysis is a tool often used in metabolomic investigations for data reduction and the correlation of spectral signatures to population identification and statistically significant separation. Although multivariate analysis has been around for many years it was primarily developed by, and for, those specializing in the field of statistics. Due to this fact collaborations are usually formed between biologists or clinicians and biostatisticians in order to bring together the two disciplines. It is important to have a basic understanding of the two fields to ensure both the analytical and the statistical tools are implemented appropriately, allowing for the proper analysis and extraction of the data. *This appendix offers some discussion points that are important to consider when using multivariate statistical tools with NMR analysis in metabolomic investigations.*

INTRODUCTION

Following NMR spectral acquisition it is important to consider the methods used to extract the relevant information. One-dimensional ^{1}H-NMR provides spectral data that contains both qualitative and quantitative information.

Resonant peaks unique to each particular metabolite are separated according to frequency, with the area under the peak corresponding to the concentration. This amount of detailed information is extremely useful for the identification and monitoring of metabolites as pathways are altered during a specific disease pathophysiology. However, because NMR spectra are full of metabolite resonant peaks it is extremely important that as much of the data as possible is extracted to ensure vital information is not lost.

Chapter II described work that ensured the information acquired by NMR and quantified by different software techniques was optimized for accuracy and precision. Software provided by Varian Inc. (Palo Alto, CA.) and Chenomx (Edmonton, AB.) allowed for the identification and quantitation of a number of metabolites with a defined and acceptable window of quantitative error. Such *a priori* methods of metabolite identification allowed for rapid qualitative and quantitative metabolic profiling of sampled biofluids. When the metabolites are known, the NMR resonant peaks can be quickly identified and quantified, and a general understanding of the biofluid and the metabolic state of the sampled subject can be quickly surmised. Both software techniques highlighted in Chapter II depend upon the identification of known resonant signatures and a pre-existing knowledge of the metabolites being quantified.

Another technique used to analyze large datasets is multivariate statistical analysis. Many different algorithms fit within this method of data analysis, but generally the purpose of this method is to take large amounts of information and reduce the dimensionality to a size that is more manageable and can easily be visualized and interpreted by the investigator. This method of analysis is mathematically intensive and highly specialized. Often this causes most researchers to form collaborations with a statistics lab or a mathematician who specializes in this form of analysis. The potential contribution of multivariate statistical analysis in metabolomic investigations is the ability to probe the data without having any prior knowledge of the spectral signatures, metabolites, or metabolic pathways that are expected to appear or to be altered.

This thesis is not the appropriate venue for an in-depth review of all multivariate methods used in metabolomic investigations. Rather, I will outline some of the considerations that investigators should be aware of when using multivariate analysis, in particular when applied towards 1D ^{1}H-NMR metabolomic studies. I will also identify some of the solutions discovered during my research.

A.1 Data Export

Primarily, the conversion of NMR data to a format convenient for multivariate analysis utilizes a binning or windowing technique. Essentially the spectra are divided into segments of equal size with a value for each specified region expressed as an integration of the spectral features (*i.e.* area under the curve for each window). Often with different software packages users are able to specify the size of the bin windows, or the number of windows used over the whole spectrum.

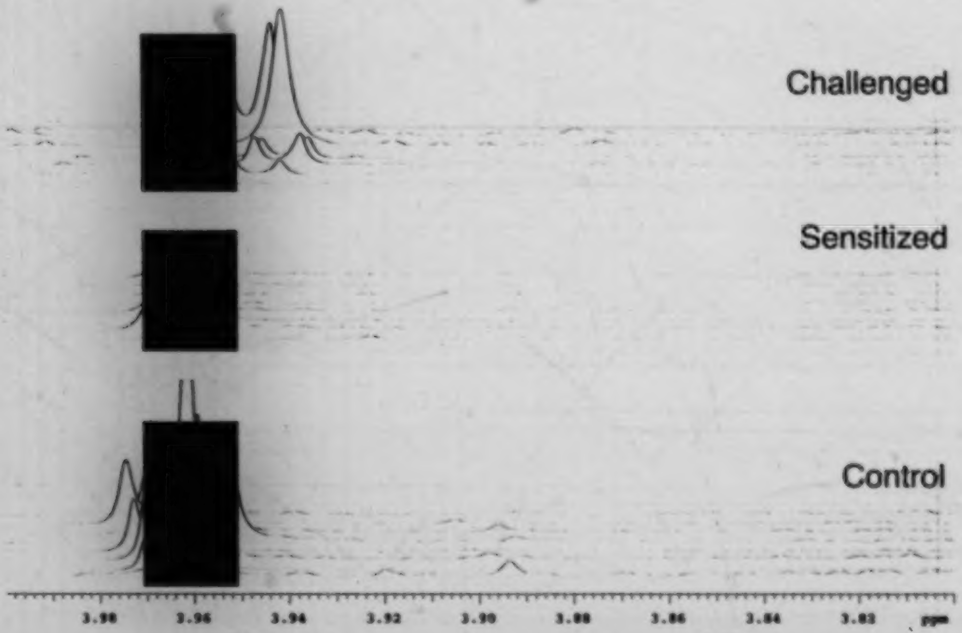
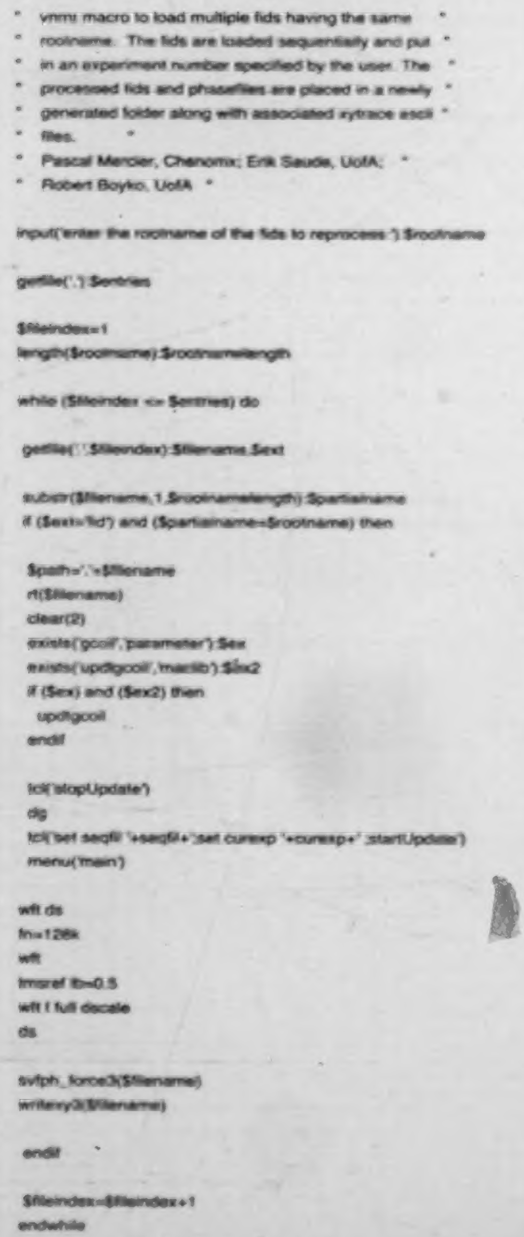


Figure A-1. **Highlighted regions of stacked guinea pig spectra**. Stacked sections of 1D ^{1}H-NMR urine spectra collected from a guinea pig model of asthma. The highlighted region represents a typical binned region for multivariate analysis. Although the resonant peaks are similar in each guinea pig group the binning window almost completely misses the peaks in the challenged group, while sensitized and control binned regions contain the resonant peaks.

The spectra collected for the majority of this thesis contained 65K complex data points (65536). The Varian software used in my research contains a macro called 'xytrace' which divides the spectrum into a number of points specified by the data digitization (number of points, np). Statisticians find it helpful to consider each data point representing a compound or metabolite and the amplitude as the rough estimation of the concentration.

The exporting of the spectra in XY format, the re-saving of the spectra for Chenomx analysis, and the re-naming of all the files was automated using a macro created by Pascal Mercier, Erik Saude, and Robert Boyko (autoprocess). Spectra were exported into individual ASCII files using the software suppliers' macro 'xytrace' and the software package VNMR 6.1C (Varian Inc, Palo Alto, CA). The spectra were exported into an ASCII file, which contained two columns: data point in Hz. (X) and the corresponding amplitude (Y). Spectra were exported into 32k data points (32768). Figure A-2 shows the macro autoprocess and the spectral processing that took place prior to export. Figure A-3 shows the modified VNMR svfph macro used by 'autoprocess' ('writexy' modifications not shown).



```
"   vnmr macro to load multiple fids having the same      "
"   rootname.  The fids are loaded sequentially and put  "
"   in an experiment number specified by the user.  The   "
"   processed fids and phasefiles are placed in a newly  "
"   generated folder along with associated xytrace ascii "
"   files.         "
"   Pascal Mercier, Chenomx; Erik Saude, UofA;    "
"   Robert Boyko, UofA   "

input('enter the rootname of the fids to reprocess:'):$rootname

getfile('.'):$entries

$fileindex=1
length($rootname):$rootnamelength

while ($fileindex <= $entries) do

 getfile('.',$fileindex):$filename,$ext

 substr($filename,1,$rootnamelength):$partialname
 if ($ext='fid') and ($partialname=$rootname) then

  $path='.'+$filename
  rt($filename)
  clear(2)
  exists('gcoil','parameter'):$ex
  exists('updtgcoil','maclib'):$ex2
  if ($ex) and ($ex2) then
    updtgcoil
  endif

  tcl('stopUpdate')
  dg
  tcl('set seqfil '+seqfil+';set curexp '+curexp+' ;startUpdate')
  menu('main')

wft ds
fn=128k
wft
tmsref lb=0.5
wft f full dscale
ds

svfph_force3($filename)
writexy3($filename)

 endif

 $fileindex=$fileindex+1
endwhile
```

Figure A-2. **Macro for autoprocessing and exporting multiple spectra**. Autoprocess macro created to automatically enter a large folder of NMR data, create a new folder in which to store the autoprocessed spectra exported into XY ASCII files and re-saved with the datdir file for later Chenomx analysis.

```
"@(#)svf 16.1 10/25/02 Copyright (c) 1991-1996 Varian Assoc.,Inc. All Rights Reserved."
" svfph - save data set and datdir for third party processing"
"modified by Erik Saude and Robert Boyko,NANUC,  2005

shell('\rm -rf Autoprocess'):$dum
shell('mkdir Autoprocess'):$dum
$outdir = 'Autoprocess' + '/' + $1

if ($# > 0.5) then
  $args = 'SVF(\''+$outdir
  $i = 1
  while ($i<$#) do
   $i = $i + 1
   $args = $args+'\',\''+$($i)
  endwhile
  $args = $args+'\',\'force\')'
  dg
  exec($args)
  $cmd='cp -fr ' + curexp + '/plot.def ' + $outdir + '.fid 2> /dev/null;cat'
  $cmd1='cp -fr ' + curexp + '/process.def ' + $outdir + '.fid 2> /dev/null;cat'
  $cmd2='cp -fr ' + curexp + '/fcdata ' + $outdir + '.fid 2> /dev/null;cat'
  $cmd3='cp -fr ' + curexp + '/fcrunlog ' + $outdir + '.fid 2> /dev/null;cat'
  $cmd4='cp -fr ' + curexp + '/datdir ' + $outdir + '.fid'
else
  $filename = ''
  SVF:$filename
  $cmd='cp -fr ' + curexp + '/plot.def ' + $filename + ' 2> /dev/null;cat'
  $cmd1='cp -fr ' + curexp + '/process.def ' + $filename + ' 2> /dev/null;cat'
  $cmd2='cp -fr ' + curexp + '/fcdata ' + $filename + ' 2> /dev/null;cat'
  $cmd3='cp -fr ' + curexp + '/fcrunlog ' + $filename + ' 2> /dev/null;cat'
  $cmd4='cp -fr ' + curexp + '/datdir ' + $filename
endif
ds
flush
shell($cmd):$dum
shell($cmd1):$dum
shell($cmd2):$dum
shell($cmd3):$dum
shell($cmd4):$dum
```

Figure A-3. **Macro for processing and exporting multiple spectra.** VNMR macro svfph_force modified by Erik Saude and Robert Boyko to allow for automated multiple-spectrum export for third-party analysis.

Generally, multivariate algorithms take the data from all the spectra and perform a diagonalization of the matrix in order to identify the appropriate eigenvalues and eigenfunctions. The eigenfunctions correlate to unique NMR spectral features that distinguish the different populations being studied. Typically, the signatures or classifications used to identify and separate the different populations are a linear combination and only together do they allow for statistically significant separation. The ability to separate different populations is often the final goal of more clinically focused studies.

A.2 Binning Bifurcation of Resonance Signatures

Binning of NMR spectra usually involves the segregation of each spectrum into bin windows that have a set width and periodicity. Since each window is of a specified size (*e.g.* 0.01 ppm) there is an assumption that the spectral signatures captured within will be similar between different spectra. The difficulty is that window placement and width is usually specified over the full spectrum and often a single peak can traverse different windows. With one peak bifurcated by a window border a certain amount of information regarding that metabolite is lost.

Work in collaboration with the Dr. Shah laboratory in the Faculty of Engineering has allowed for the creations of 'floating' bin windows that have variable widths. The computer program has a minimization algorithm that allows for some minor freedom in placement of the window borders. In the event that the window falls within the middle of a peak, the algorithm tries to find a nearby minimum in which to place the final border. This will reduce the number of resonant peaks bifurcated by window borders, although it also creates some discrepancies in the number of bins per spectrum and bin regions when comparing different spectra during multivariate analysis.

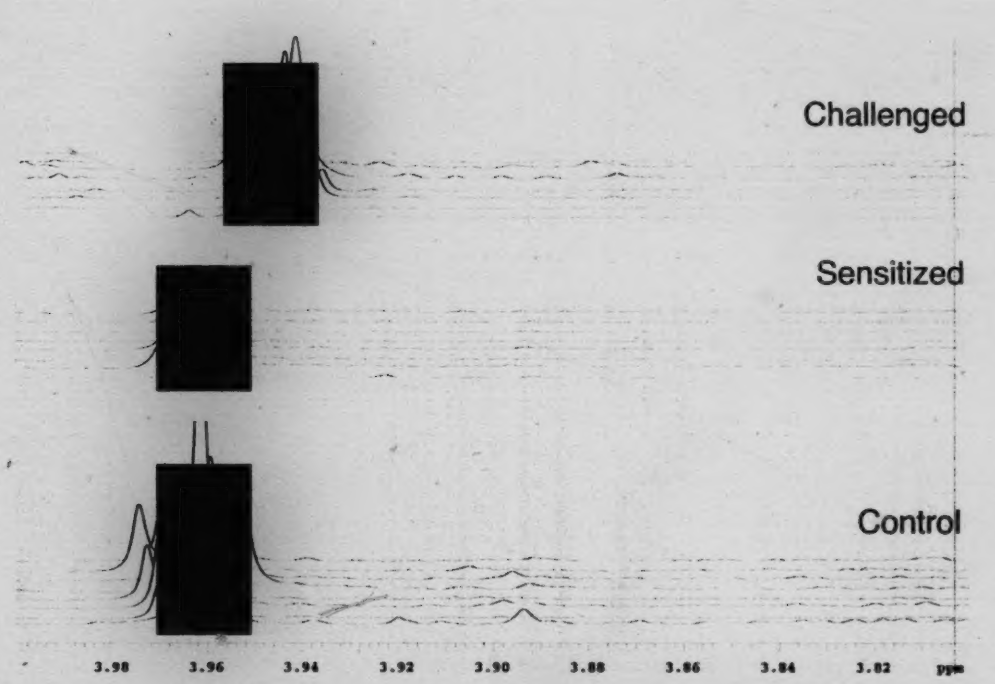


Figure A-4. **Stacked spectra with shifted highlighted region**. Stacked sections of 1D ^{1}H-NMR urine spectra collected from a guinea pig model of asthma. The highlighted region represents a binned region for multivariate analysis that is allowed to shift slightly left and right in order to limit the bifurcation of resonant peaks. This method helps to capture the resonant peaks in each guinea pig group.

A.3 Resonant Shifts

Resonant peaks within each window can differ between spectra due to changes in chemical shift. This difference may lead a researcher to claim absence of a signature or metabolite in one sample, when in fact it is simply in an adjacent bin. Errors in resonance capturing through automated binning is typically the result of minor fluctuations in metabolite chemical shifts as a result of pH, sample temperature during acquisition, or differing salts concentrations in the sample. These factors can cause changes in chemical shift significant enough to move resonant peaks from one bin window to another. To complicate this further the direction and degree of the change in chemical shift is unique to each compound and spin system so a generalized correction function is difficult to implement.

As a result of these binning errors samples are prepared and spectra are acquired following established standard operating procedures (SOP's). As an example, all of the urine samples for the purpose of this thesis were pH'd to 7.0 +/- 0.1, spectra were acquired at 25°C with a specified equilibration time before acquisition, and all spectra were shimmed to a set linewidth. These measures helped reduce the fluctuation of chemical shifts when compared over hundreds of samples.

A.4 Principle Components

During multivariate analysis the algorithms pick out spectral features that characterize or distinguish the different populations under investigation.

During PCA the first two or three principal components in the analysis describe roughly 98% of the variation between the two populations being studied. However, this is misleading since the first principal component is often the residual water peak in the NMR spectrum and the second principal component is the urea peak. Often neither of these peaks contain relevant information for the disease being studied and spectral fluctuations can occur as a result of NMR acquisition and not necessarily the result of actual physiological differences in the populations. As well, sometimes compounds identified in the sample are the result of sample preparation or handling. Chemicals such as DSS and imidazole are often added to urine samples to serve as a reference peak and a pH indicator, respectively. The position and concentration of these chemicals can vary slightly (*e.g.* pipetting error, binding, resonant shifts) and may alter statistical analysis.

It is important to identify and understand the multivariate techniques being used, what the algorithms are identifying as statistically significant differences, and the final relevance or biological interpretation of the analysis. Many journals accept studies and analysis that simply demonstrates population separation and publish figures showing separation based upon a blind multivariate analysis approach with no spectral or metabolic interpretation of the data. Without a continuation of the analysis the reason behind the separation is not known and is equally likely to be the result of error than a novel discovery.

A.5 Bin Normalization

Basic multivariate analysis for general population separation based on exported NMR data is not ideal. In order to ensure the analysis is performed properly the initial data acquisition should be optimized and the statistical analysis should also be understood before any conclusions are drawn. In addition, studies typically lack urine normalization of the bins before analysis. Similar to absolute metabolite concentrations when determined by software packages like Chenomx (see Chapter IV), the spectral signatures should be normalized to creatinine prior to multivariate analysis. As discussed in Chapter IV, metabolite concentrations can fluctuate as a result of subject hydration and may not represent the particular physiology being studied. The accepted practice of the general scientific community is to normalize metabolite concentrations to the creatinine concentration of the sample.

For the multivariate analysis performed in the studies for this thesis an in-house macro had to be written to ensure all urine sample spectral bins were normalized. This was executed by first identifying and restricting the creatinine bin to the single metabolite resonance and determining the bin volume. The bin volume for creatinine was then used to normalize remaining spectral regions in an attempt to remove hydration variation in urine metabolite quantitation. This method tends not to be ideal because of the differences in resonance volume, as well as the number of protons contributing to the resonance. However, this method is superior to other multivariate normalization methods such mean-centering or normalizing to total spectral area.

A.6 Data Conversion

It is important to understand the translation of data from one format to another, particularly when exchanging data with collaborators that use different formats (*e.g.* spectra in ppm, and exported data expressed as data point). Multivariate analysis computer programs and algorithms use raw binary formatted data (*e.g.* data points, or number of points, np) while the spectroscopist prefers to use spectral terms such as hertz or ppm. Sometimes individuals or programs have to automatically interconvert the information between data points, hertz, and ppm data. The concern is that often neither individual has enough understanding of the other discipline and as such errors often occur as a result. For instance to achieve ppm from data expressed in hertz, a program might automatically divide hertz by the spectrometer strength, 600MHz, when a more accurate resonant frequency would be 599.978Hz. This inaccuracy is significant enough to shift binning windows and the statistically significant spectral signatures to new regions causing errors in the final correlation of spectral characters to actual resonant peaks.

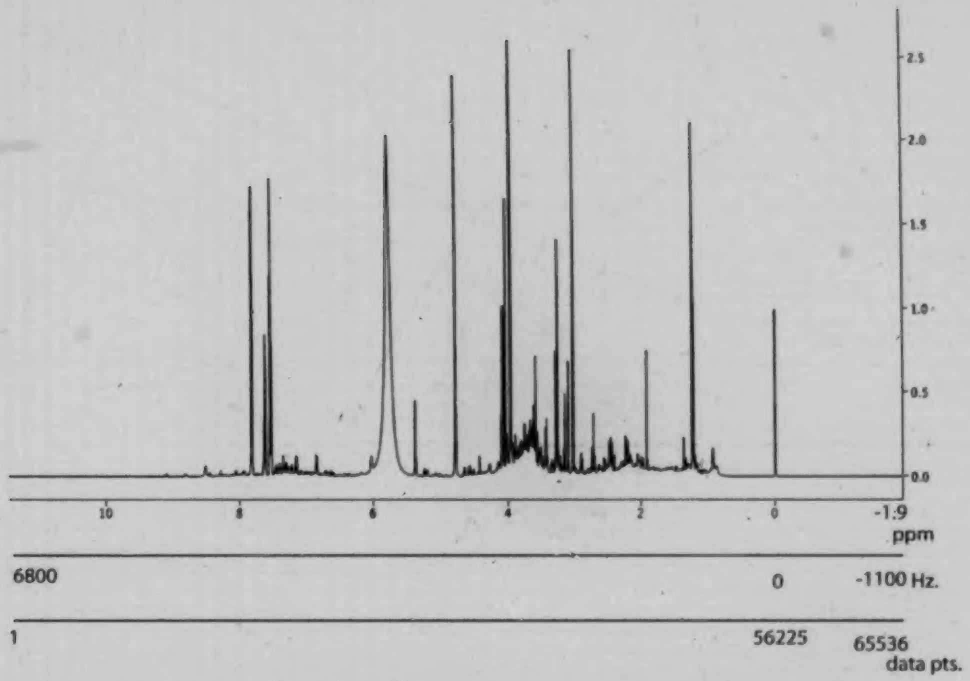


Figure A-5. **Spectral axis defined as ppm, Hz, and data points**. A typical guinea pig 1D 1H-NMR spectrum with the widely accepted 'ppm' frequency scale, Hz scale, and exported data point axes. This figure demonstrates the errors that might arise when changing from different reference axis formats.

A.7 Statistical Analysis

There are many methods for multivariate statistical analysis; principal component analysis (PCA), neural networks, relative distance planes (RDP), partial least square (PLS), etc. Although a review of each method would be far too lengthy for the purpose of this thesis appendix, it is important to mention that the questions posed by the researcher will ultimately dictate which multivariate method is most appropriate. As well, researchers should be reminded that often some preparation of the data is required prior to analysis (see above).

Researchers should be cognizant of the general outcomes they require following multivariate analysis. Often analysis will allow for separation of the different populations in the study, however, if an investigator expects to identify specific compounds and metabolic pathways additional work is required. As an additional reminder, the principal components or spectral signatures identified by multivariate analysis are 'diagnostic' when used in combination so no single spectral region or metabolite will be able to separate the populations if the analysis is repeated.

A.8 Statistics

When discussing multivariate analysis and statistics it is appropriate to address the issue that arose in Chapter IV, the normals study, regarding proper statistical analysis of human metabolic data. Inherent in the methods of statistical analysis commonly used in biology (*e.g.* ANOVA, student t-test) are assumptions regarding variance. For example the ANOVA (analysis of variance) tests the null hypothesis that group means do not differ. It is not a test of differences in variances, but rather assumes relative homogeneity of variance. Therefore, an assumption remains that the groups formed by the independent variables are relatively equal in size and have similar variances on the dependent variable ('homogeneity of variances'). As in regression analysis, the ANOVA is a parametric procedure, which assumes multivariate normality (*i.e.* the dependent has a normal distribution for each value category of the independent). Following the normal study and the large degrees of variation observed I have some hesitations regarding the currently accepted methods of reporting statistical significance in many of the scientific journals when dealing with metabolomic data.

I am currently conducting a large literature review of metabolomic and metabolomic research articles and their methods of statistical analysis. I believe efforts should be made to improve and standardize statistical and reporting methods to ensure final conclusions are as accurate as possible. I recently attended a scientific talk that discussed efforts currently underway by the Metabolomics Society to do just this; standardize statistical analysis and reporting methods.

APPENDIX B

Novel biomarker identification

OVERVIEW

Metabolomics attempts to identify metabolites in sampled biofluids and correlate qualitative and quantitative differences or fluctuations to the disease being studied. Typically metabolites are commonly occurring and may indicate general biological distress (*e.g.* lactic acid). In some instances novel spectral signatures are resolved and further biochemical experiments are required in order to identify the chemical structure of the new biomarker and pathophysiological pathway of origin. *This appendix depicts some experiences I have had regarding the identification of some novel spectral signatures, their chemical description, and the biological relevance.*

INTRODUCTION

Metabolomic investigations involve the collection of large amounts of data (*e.g.* NMR spectra) through biofluids collected from different test subjects experiencing different physiological states. Through the use of multivariate statistical analysis researchers may circumvent the requirement for any prior knowledge regarding known metabolites in a studied biofluid, as well as the possible changes during a disease process. Through multivariate analysis different populations are distinguished based on the identification of unique spectral signatures. The advantage of multivariate analysis, for metabolomic studies, is the often non-supervised nature of the method. This minimizes the possibility of over fitting by the researcher. Through multivariate analysis researchers hope to receive identifiers from the algorithm that represent unique spectral characteristics that can potentially be novel biomarkers.

Appendix A discussed some of the considerations when performing multivariate analysis on NMR data for metabolomic studies. Chapters III and V have discussed the importance of correlating distinctive NMR signatures to a biochemical or physiological understanding of the biological system. Once multivariate analysis has reduced the dimensionality of the analysis to a few key spectral characteristics additional NMR and biochemical experiments can be performed to help with the identification of the metabolite. This appendix provides examples of analytical techniques and experiments I have used in order to identify novel biomarkers that have arisen following metabolomic investigation of biofluids.

B.1 1D NMR

Asthma investigations using urine analysis allowed for a much easier sampling of a larger population than earlier studies that utilized sputum collection. Initial asthma urine analysis attempted to find the oxidized tyrosine residues bromo- and chlorotyrosine that had been identified earlier in the sputum.

However, due to dilution effects the modified tyrosine residues were in much too low of a concentration for NMR detection. Early analysis continued to focus on the aromatic region of the spectrum (6 – 10ppm) and returned the initial discovery of a compound never before identified by my laboratory. The initial finding was of a doublet/doublet resonance pattern (7.1ppm) similar to the oxidized tyrosine residues, but at a different chemical shift.

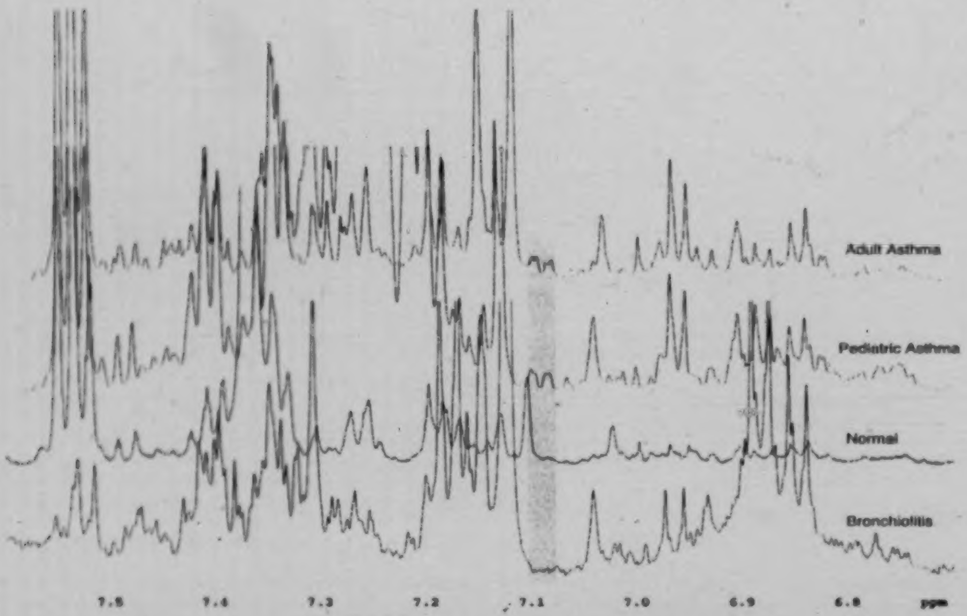


Figure B-1. **1D ^{1}H NMR spectra of the unknown compound**. Stacked plot of four 1D ^{1}H-NMR spectral aromatic regions of urine from adult asthma, pediatric asthma, normal, and bronchiolitis patients. The doublet-doublet at 7.1 ppm is highlighted as a unique spectral signature from an unknown metabolite.

The identification of an unknown compound represented a new task for metabolomic investigations, a unique spectral signature with an unknown origin, but with possible biological relevance to a unique clinical state. The first step was to identify correlated resonant peaks. NMR spectra offers structural and biochemical information through interpretation of the different resonant frequencies, multiplicities, and couplings a compound can produce.

To identify correlated resonant peaks a one-dimensional NOE experiment was designed where the magnetization was allowed to transfer between coupled spins. This allowed for the identification of correlated resonance systems in the one-dimensional spectrum. With help of Dr. Ryan Mckay, we created a one-dimensional ^{1}H NOE experiment with an initial shaped pulse that excited the doublet-doublet spin system that resonated at 7.1ppm and allowed the magnetization to exchange with nearby spins.

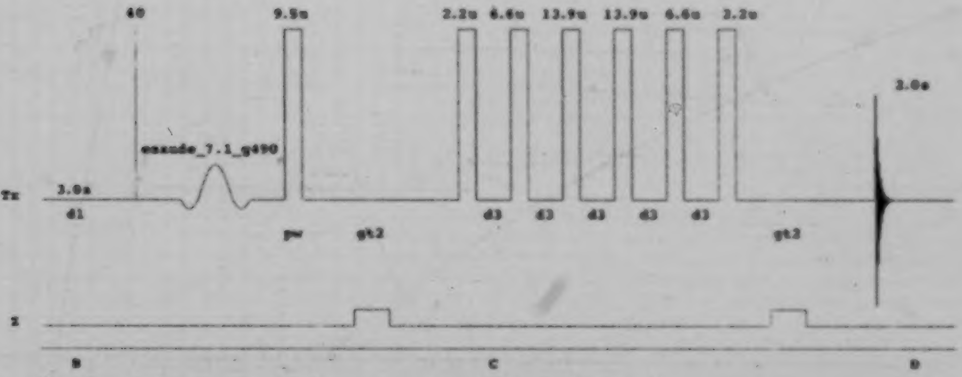


Figure B-2. **Schematic of the 1D-NOE pulse sequence**. The shaped pulse at the beginning of the pulse sequence was optimized for yellow 7.1 (doublet-doublet at 7.1 ppm).

The 1D NOE experiment showed some signal exchange that helped to identify possible resonant peaks that originated from the same compound as '7.1'. For ease of identification and communication the different spin systems were colour coded (see Figure B-3). The yellow resonant peaks are those that were excited by the shaped pulse and originated from the same metabolite. The orange peaks represent a correlated coupling system, possibly from another molecule. The identical coupling structure, but slightly shifted frequencies of the orange compound when compared to the yellow-7.1 compound, opens the possibility that the orange compound was either similar or the yellow, but might be in another conformation (Figure B-3). The blue resonance peaks are from tyrosine and were excited by the original shaped pulse focused around 7.1ppm.

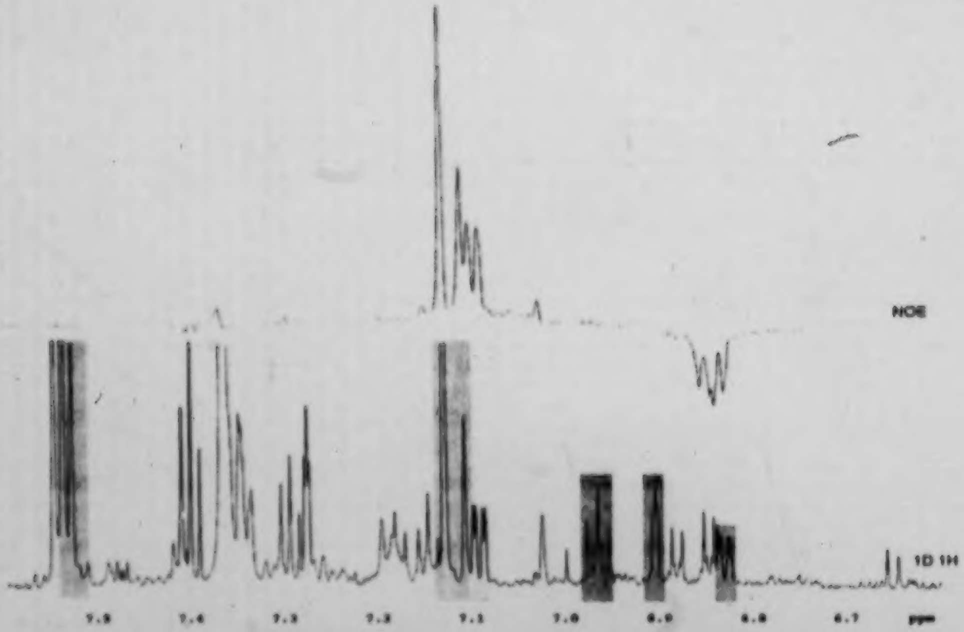


Figure B-3. **1D-NOE and 1D ^{1}H spectra of asthma urine.** Stacked plot of 1D NOE and presat pulse sequences. Selective excitation of the doublet/doublet of yellow 7.1 by the NOE experiment helped to identify correlated resonances. Correlated peaks were confirmed by 2D experiments, see Figure D-4.

B.2 Multidimensional NMR

Following the initial 1D NOE experiment, some two-dimensional experiments were fun to help support and perhaps identify additional resonant peaks originating from the 'yellow and orange' metabolites. Some optimization of the 2D data acquisition was required due the NMR behaviour of the small metabolites (see Figure B-4). The experiments included COSY (correlation spectroscopy), DQF-COSY (double quantum filtered correlation spectroscopy), NOESY (nuclear Overhauser effect spectroscopy), ROESY (rotational NOE spectroscopy), and DISPI (decoupling in the presence of scalar interactions).

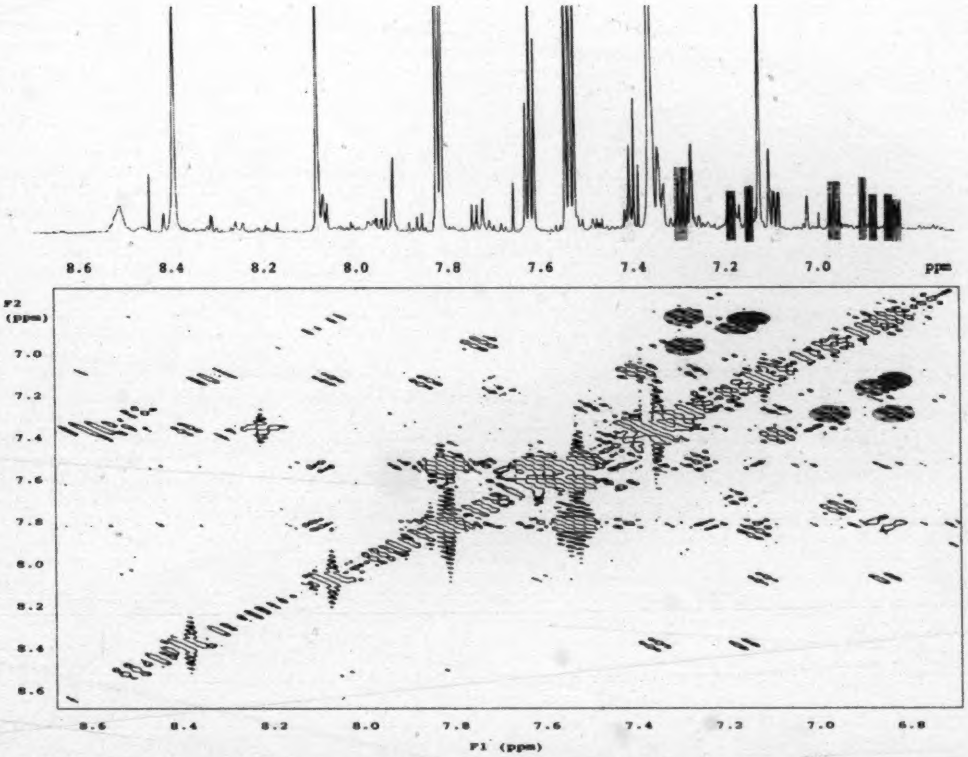


Figure B-4. **2D and 1D NMR spectra of the unknown metabolite in human urine.** 2D NOESY spectrum with a 1D ^{1}H-NMR aligned on top in order to identify correlated peaks in the two experiments (both spectra from one asthma patient's urine). Correlated peaks in the 2D and 1D experiments are coloured accordingly. The unknown resonance yellow 7.1 has a similar pattern that is duplicated and shifted up-field, orange 6.8.

The information provided by the 1D and 2D experiments indicated that the metabolite was likely a bi-substituted benzene ring. Although not shown in the figure above, some weak couplings were observed in the 2D experiments between the benzene ring and likely a CH2 group. This raised the possibility that the metabolite might be similar to hippurate, but with a substituted benzene ring.

Chemical standards were purchased and analyzed by NMR in order to identify similar coupling patterns and chemical shifts as the unknown metabolite.

B.3 Chemical Standard

The strength of NMR is easily observed in the stack plot of the different spectra from the chemical standards (Figure B-5). Each chemical has a relatively similar pattern of resonant peaks and couplings; however, chemical shifts are unique to each compound. The different electronegativity of each substituent alters the resonant frequency of the nearby spin systems generating a unique spectral pattern. Information from multidimensional NMR and the comparison with chemical standards suggests that a substituted hippurate, for instance 4-aminohippurate, is likely the unidentified metabolite yellow 7.1 identified in the urine. Continuing work with other techniques such as MS will help confirm the presence of this compound in the urine.

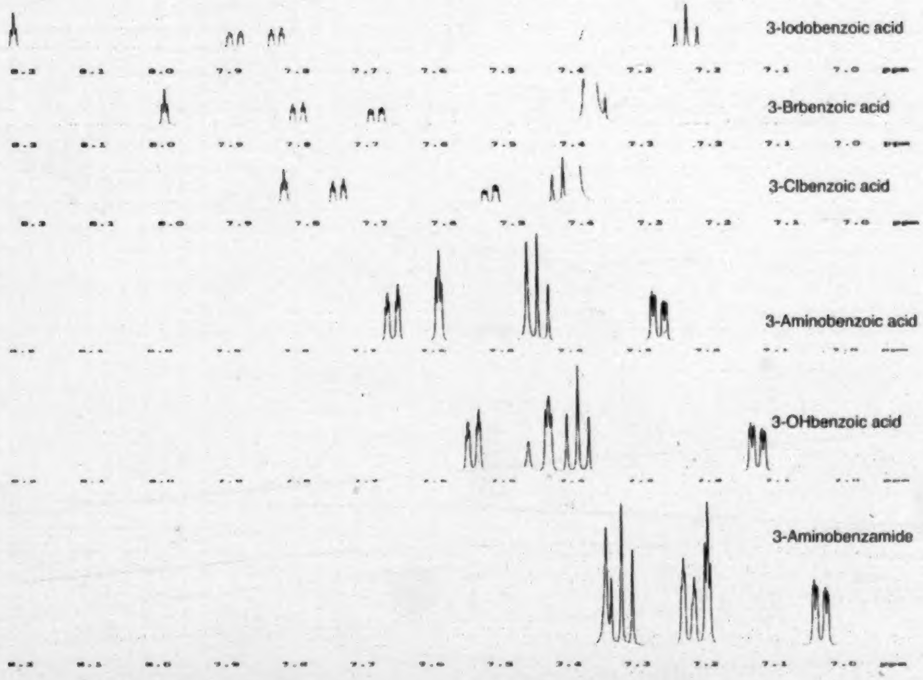


Figure B-5. **1D ^{1}H-NMR spectra of chemical standards**. Stacked aromatic regions from pure chemical standards thought to be similar to yellow 7.1 analyzed by 1D ^{1}H-NMR. Though the chemicals are structurally similar and produce similar resonant peaks, this figure demonstrates significant chemical shift fluctuations between different chemicals.

B.4 Double Quantum Filtered 1D-NMR

Earlier asthma work for this thesis identified the possibility of brominated tyrosine residues in the sputum as biomarkers for the disease. My supervisor, Dr. Sykes, and I developed a double quantum filtered 1D ^{1}H-NMR pulse sequence that allowed for the removal of the majority of the resonant peaks in the aromatic region while leaving the singlet from 3,5-dibromotyrosine largely untouched (see Chapter V). Unfortunately, the concentration of brominated tyrosine residues in human urine is below the detection limit of NMR so the pulse sequence is no longer used; therefore, the technique will not be explained further. However, the double quantum filtered 1D ^{1}H-NMR pulse sequence provides yet another example of possible pulse sequences that may be developed in order to assist in biomarker identification.

B.5 Mass Spectrometry

In Chapter I, some disadvantages of MS were highlighted. In short, MS requires lengthy sample preparation and involves sample destruction during analysis. However, an advantage to MS analysis is the greater sensitivity when compared to NMR. By combining the strengths of both NMR and MS researchers are able to gather a more complete profile of metabolites present in urine. This thesis has already shown through the asthma metabolite investigation mentioned above, how NMR is able to identify a novel metabolite resonance and formulate a likely chemical structure. With the rough chemical formula HPLC and MS analysis of the urine is able to confirm and offer additional information regarding NMR metabolite identification.

B.6 Unique Metabolite in Guinea Pig Urine

Another unique compound was identified in the 1D ^{1}H-NMR spectra of guinea pig urine. The resonant peaks were very large and were strikingly similar to the resonant patterns for ethanol. However, the chemical shifts of the unknown compound were downfield from where ethanol is usually found.

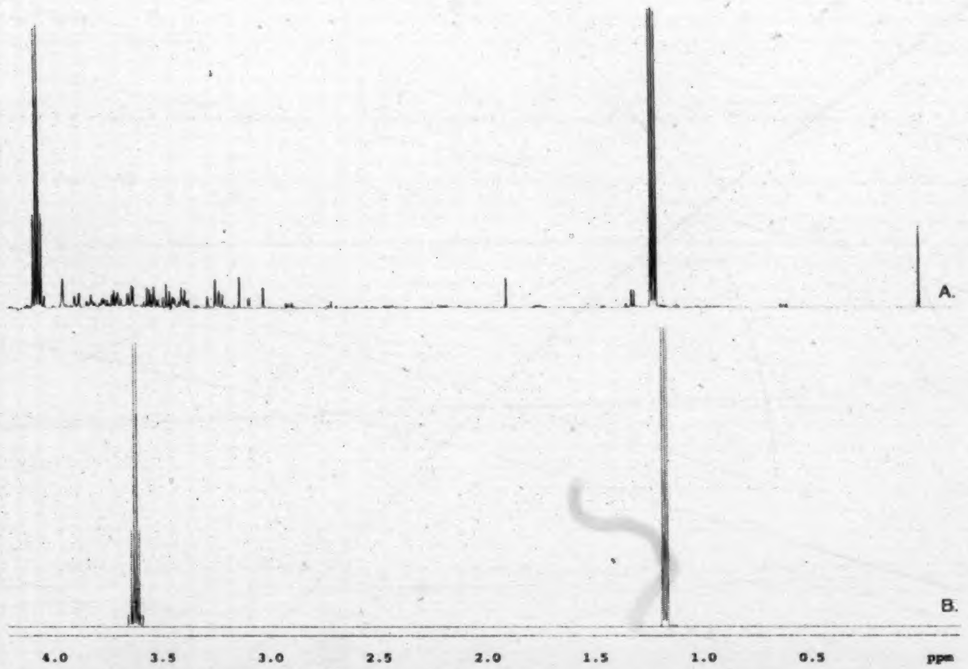


Figure B-6. **Spectral analysis of a guinea pig urine NMR spectrum**. 1D 1H-NMR spectra (0 – 4 ppm) of guinea pig urine (A), the Chenomx NMR Suite Database signature, and chemical shifts for ethanol (B). Although the couplings and relative amplitudes are similar, the difference in chemical shift between the resonances in the guinea pig urine and the database indicated that the compound in question was not ethanol.

The general chemical shift, coupling pattern, and area of integration indicated that the resonant peaks were due to a CH_2 and CH_3 spin system (4.0 and 1.2ppm, respectively). The shift in resonant frequency with respect to ethanol indicated that the substitute on the aliphatic compound was not a hydroxyl group. A lack of other resonant peaks with similar amplitude in the NMR spectrum indicated that the substitute on the unknown compound was not NMR active. The urine was submitted for HPLC and MS analysis to determine the mass of the compound. Automated analysis of the MS spectra offered a number of chemicals that could theoretically have a mass similar the compound under investigation, but none of the chemicals had any biological relevance.

During a discussion Dr. Adamko (supervisor of animal work) regarding the handling of the guinea pigs it was identified that the anesthesia used for the animals during sample collection was called Urethane, or ethyl carbamate. This compound has a molecular mass identified during MS analysis, a CH_3CH_2 group observed by NMR, and was later confirmed by NMR analysis of a pure chemical standard of urethane. Since a portion of urethane cannot be observed by NMR, MS helped identify a list of possible chemical compounds, which helped lead to the identification of the anesthesia in the guinea pig urine.

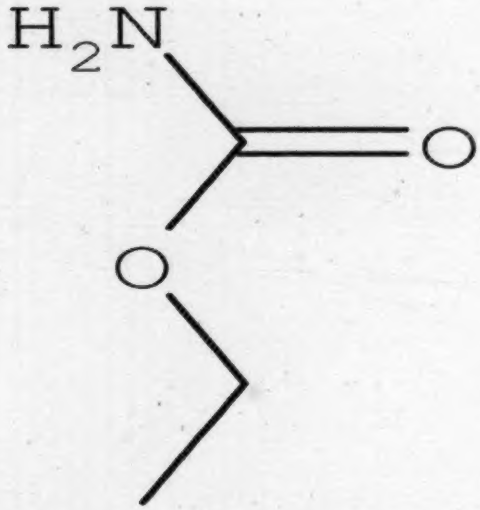


Figure B-7. **Chemical structure of the anesthetic urethane.**

This example also lends importance to the careful and thorough investigation of body fluids. Compounds identified in urine may originate from a number of sources, including; sample handling, sample collection, and animal care. Great efforts must be taken to ensure analytical analysis and final data interpretation have accounted for all experimental factors in order to limit scientific error.

CONCLUSION

This analysis offers examples of additional work performed following the identification of unique spectral signatures in an effort to reach a final biochemical understanding. This appendix raises the importance of conducting further analysis to achieve a complete biochemical understanding in order to avoid 'false diagnosis' based upon a spectral signature that is not truly diagnostic.

As well, this appendix reveals some examples that demonstrate the wealth of experiments available to analytical investigations through the tandem use of NMR and MS. By exploiting the strengths of the two techniques a fuller description of the metabolic profile is possible.

APPENDIX C

NMR Spectral Baseline

OVERVIEW

As a supplemental to Chapter II, this appendix will continue the discussion surrounding metabolite quantitation. The strength of NMR lies in the ability to analyze complex biofluids in both a qualitative and quantitative manner. Clinical and metabolomic research would be much easier if there was a single metabolite unique to every disease. The reality is that diseases often share similar metabolites and the distinction from one disorder to another is the unique combination of metabolites, as well as their respective quantity. A significant factor to spectral quantitation is the stability of the NMR baseline. In Chapter II some baseline issues were raised. I would like to further discuss some physiological and chemical considerations. *This appendix will describe work conducted to further the accuracy of NMR quantification; some of the methods included removing or stabilizing the spectral baseline through chemical, mechanical, and computational means.*

INTRODUCTION

Accurate metabolite concentration determination is extremely important in metabolomic investigations. Although some biological differences may be identified by the simple presence or absence of a key metabolite, the majority of metabolic clinical investigations include the identification of metabolites that are present in both normal and diseased states. The final manifestation of the disease is when the common metabolite reaches a concentration outside the normal range. This raises the importance of NMR-derived metabolite concentrations, from compound identification to accurate quantitation, and final physiological interpretation. The true strength of metabolomic investigations likely arise, not from the identification of a disease, which can often be determined by simpler clinical methods (*e.g.* runny nose and diarrhea; you have the flu), but in the finer identification of disease progression, severity, and response to treatment. The potential ability of metabolomics to follow minor shifts in metabolic pathways will allow for a greater understanding of biological processes, allow clinicians to follow patient improvements, and help tailor fine changes to therapeutic interventions.

Key questions emerge from Chapters II and V surrounding metabolite quantification and NMR or software-based quantitative error. NMR derived metabolite concentrations are calculated by integrating the area of a resonant peak and correlating that area with a reference peak of known concentration. Different software packages, such as those provided by Varian and Chenomx, attempt to reduce quantitative error by allowing the user to define the areas to be integrated. There are many spectroscopic sources of error that must be defined in order to fully understand how to properly and accurately determine metabolite concentrations (see Chapter II). As well, newly developed software or macro based approaches to concentration determination might allow for greater accuracy.

C.1 Classic NMR Concentration Determination

Chapter II outlines how concentrations are determined using the Varian and Chenomx software packages. Varian software allows the user to place integration windows around the peak of interest and the program returns an integrated area for that peak. A similar measurement is performed on the reference peak, often DSS, which has a known concentration. A comparison of integration areas, number of protons, and the known concentration of DSS allows for the calculation of the spectral metabolite concentrations. The Chenomx software performs a similar task to the Varian software, however, the calculations as well as some correction factors are included, in a black-box fashion, and the user is just given the final concentration. The Chenomx software performs a least-squares fit of the spectral signature to a database spectrum and calculates the concentration accordingly (see Figure IV-1).

Regular NMR practice stresses the importance of spectral quality and the impact it may have on quantitation (addressed in Chapter II). Although I will not discuss these issues further now, they were mentioned in Chapter II. It is important to remember that the general spectroscopic factors tuning and matching, spectrometer shimming, and post-acquisition processing of phasing and baseline correction affect quantitative accuracy. Figure C-1 shows the integration of DSS and citrate by Varian software for one random asthma urine spectra.

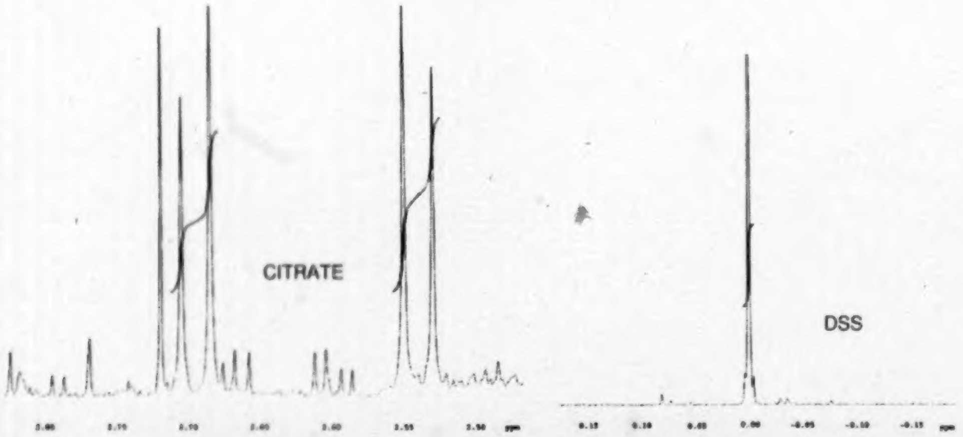


Figure C-1. **Quantitation of resonant peaks by integration**. 1D ^{1}H-NMR spectral regions for the metabolite citrate and the NMR reference compound DSS (600 MHz). Red lines indicate integration regions used for concentration calculations.

The calculation of metabolite concentrations require the identification of specific peaks and the integration of the signature region. The strength, and yet also the challenge, of high-resolution 1D ^{1}H-NMR is the ability to simultaneously visualize numerous compounds at once. In a metabolomic investigation of cultured cells, tissue, or sampled biofluid the spectral richness is a challenge. Resonance overlap is compounded with other issues of quantitation, such as baseline, and is discussed below.

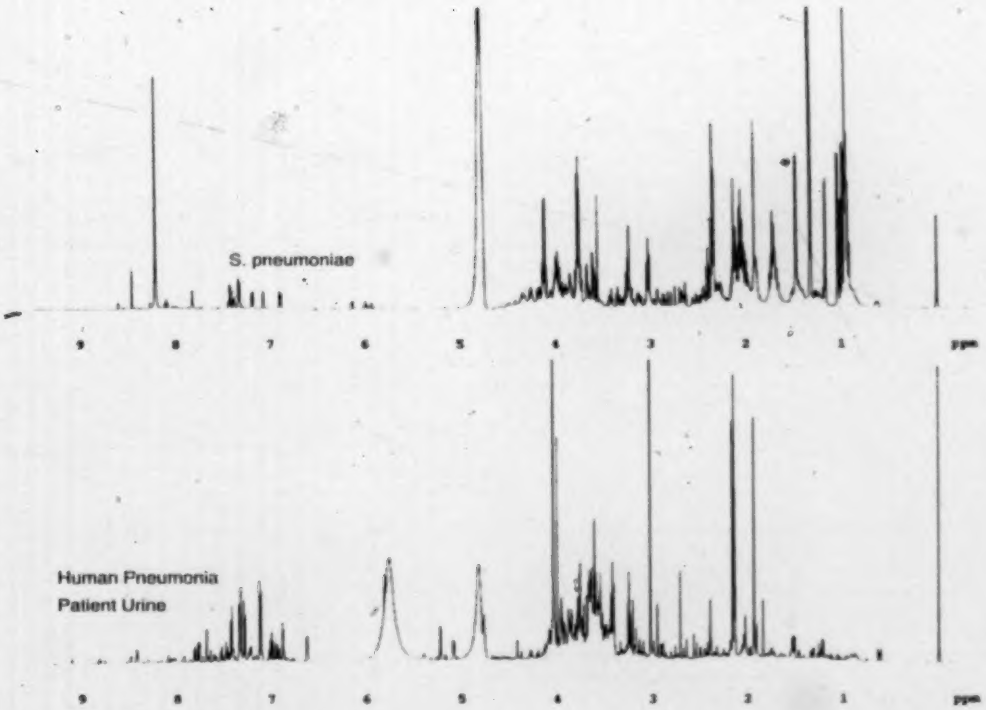


Figure C-2. **Spectral richness and resonance overlap**. 1D ^{1}H-NMR spectrum of pure cultured *S. pneumonia* (sonicated) and human pneumonia patient urine. Spectra demonstrate metabolite richness, unique signature profiles, and baseline issues.

C.2 ^{13}Carbon

The spectral area chosen to be included for integration is arbitrarily set by the user and has a large impact on quantitative accuracy. As an example, the referenced DSS singlet is in fact split by naturally abundant ^{13}C and ^{29}Si on the DSS molecule. The 29Si satellites are very close to the larger DSS singlet and are usually included in the integration. However, ^{13}C satellites are roughly +/- 0.1 ppm away from the DSS singlet and are easily excluded from the integration. The natural abundance of ^{13}C is roughly 1%, so the exclusion of ^{13}C satellites from DSS integration accounts for a drop to 99% quantitative accuracy. If the integration is limited to the points of intersect on the DSS peak from the tops of the ^{13}C satellites the integration is reduced to 96%. Remembering that each ^{13}C satellite peak is only about 0.55% of the central DSS singlet, the reduction of the integration window has an extremely significant effect on quantitation accuracy.

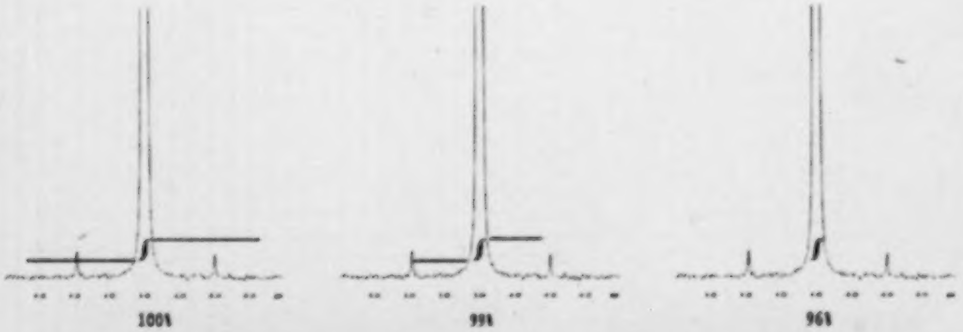


Figure C-3. **Spectral integration: effect of window size**. Spectral region of referenced DSS singlet (1D 1H-NMR) integrated with three different integration windows. Demonstrates significant loss of integrated peak with minor reduction in integration window.

It is extremely important that the size of the integration window incorporates as much of the resonant peak as possible to ensure the calculated concentration is as accurate as possible.

C.3 Peak Overlap

Although the spectral richness of human urine is a strength, when it comes to metabolite quantitation peak overlap is a significant concern. Figures C-1 and C-2 demonstrate resonance integration of separated peaks with relatively little overlap. As discussed above even the slightest reduction in the size of the integration window greatly impacts the concentration calculations. The user is forced to visually optimize the exclusion of surrounding peaks while incorporating as much of the desired resonant peak as possible.

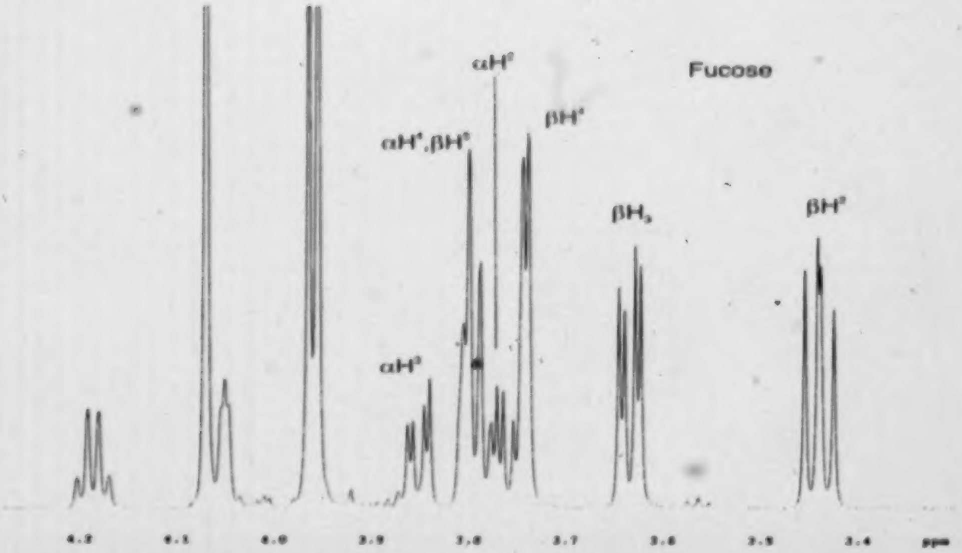


Figure C-4. **Resonance overlap in a 1D ^{1}H-NMR spectrum**. Selected region of a 1D 1H-NMR spectrum of pure chemical compounds, including the carbohydrate fucose. Resonant peaks originating from fucose protons are labeled accordingly and demonstrate spectral complexity such as multiplicity and peak overlap.

The example of spectral overlap in Figure C-3 is of a chemically created NMR sample with the sugar fucose. Clearly some resonant peaks cannot be used for concentration determination due to spectral overlap. The integration of resonant peaks such as αH4 and βH5 is impossible due to the contribution of spectral area from both spin systems. Resonant peaks that clearly interfere with each other must be excluded during concentration determination and only those signatures that can be integrated exclusively should be considered.

C.4 Macromolecules

Biofluids are collected for metabolomic investigations due to their rich composition of metabolites. However, other compounds are also present that may interfere with metabolites directly, or may interfere with the spectral resolution and quantitation. Large proteins and lipids are present in biofluids such as blood, and to a lesser extent urine. These macromolecules can interfere by directly binding to metabolites thereby interfering with their spectral resolution (*e.g.* resonance broadening or removal from the spectrum.

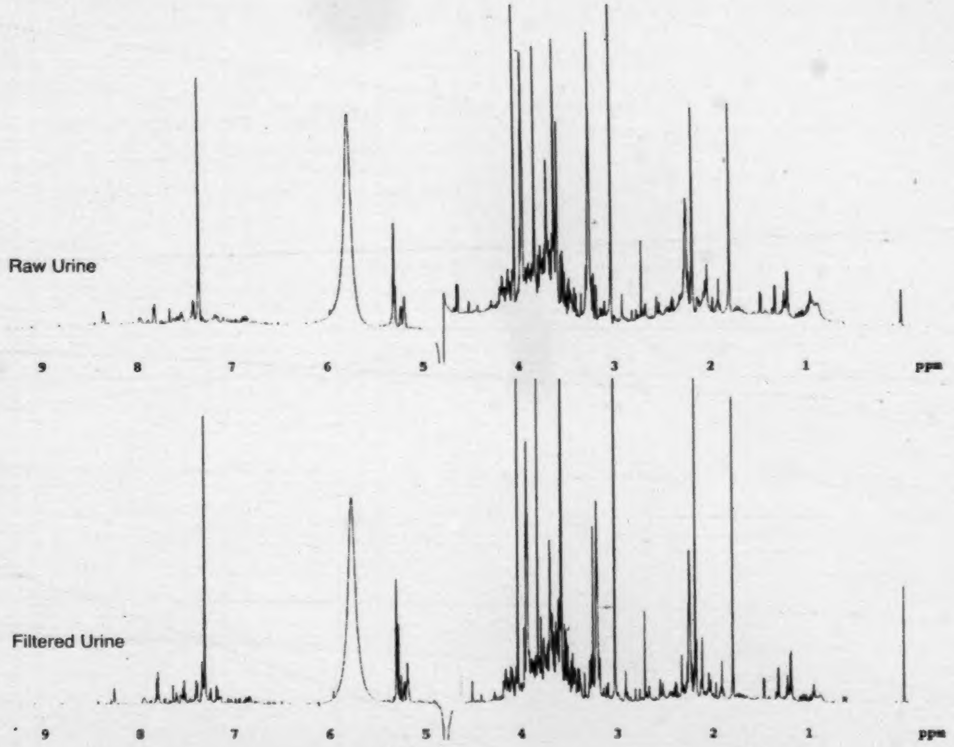


Figure C-5. **Sample filtration improves baseline**. Urine sampled from a human pneumonia patient and analyzed by 1D ^{1}H-NMR. Spectra demonstrate significant spectral improvement in baseline and metabolite resolution following urine filtration.

The urine in Figure C-4 was taken from a bacterial pneumonia patient acquired, and processed according to the methods outlined in Chapter II. Another aliquot of the same urine was then filtered (see methods section of Chapter III) and re-acquired. There was a significant change in spectral baseline after filtering the urine of macromolecules. General baseline features around 7-8 ppm, 3-4 ppm, and 0-2 ppm are greatly affected following urine filtration. As an example, the DSS singlet at 0 ppm is of the same concentration in both samples. In the filtered sample the DSS has a much greater amplitude and the resonance is no longer broadened. Similar effects are seen in other resonant peaks in the spectrum as a result of binding and interference by macromolecules.

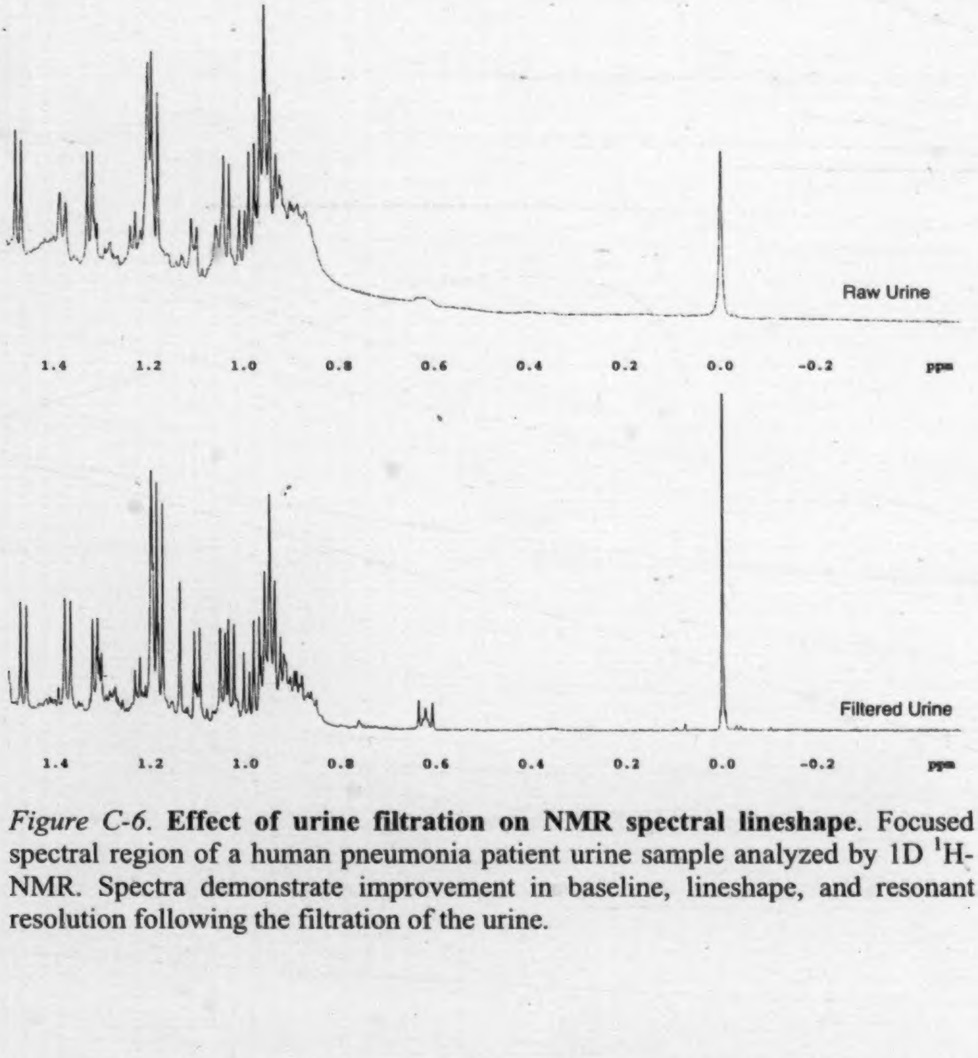


Figure C-6. **Effect of urine filtration on NMR spectral lineshape**. Focused spectral region of a human pneumonia patient urine sample analyzed by 1D ^{1}H-NMR. Spectra demonstrate improvement in baseline, lineshape, and resonant resolution following the filtration of the urine.

Similarly, Figure C-5 contains two spectrum taken before and after filtering a urine sample donated from a bacterial pneumonia patient. NMR acquisition and post-processing was identical, however, there remains a clear difference in the spectral baseline and lineshapes of the resonant peaks. The DSS singlet at 0 ppm in the raw urine had a linewidth of 3.43 Hz at 50%, 26.96 Hz at 0.55%, and 37.24 Hz at 0.11% peak height. In the filtered urine the DSS singlet of similar concentration had a linewidth of 0.66 Hz at 50%, 8.44 Hz at 0.55%, and 10.47 Hz at 0.11% peak height. In addition, the DSS resonance at 0.6 ppm was unobservable in the raw urine, except for a generalized 'lump'; while the filtered urine produced a clearly resolvable DSS resonance at 0.6 ppm.

Macromolecules have a significant impact on the ability to resolve and quantitate metabolite resonances. Investigators must be aware of the impact macromolecules can have on the identification and the accuracy of concentration calculations, and take appropriate methodological actions to correct or ameliorate such effects.

C.5 Baseline Quantitation

As mentioned above metabolite concentrations are determined by integrating the area within a resonance and within a user-defined window. It is important to consider that the baseline, as well as any distortions or elevations of the baseline are incorporated into the integration calculations. A question that arose during this study was "To what degree does the baseline affect quantitative accuracy?" "Does the resonance peak sit atop the baseline, or is the broadened baseline simply surrounding the peak and essentially not affecting the peak?" In

Figure C-6 the DSS singlet is atop a stable baseline so any integration will be representative of the DSS concentration, but the resonances from 0.6 – 1.5 ppm are amongst a broadened baseline. Should the peaks be lowered because they are atop a false baseline?

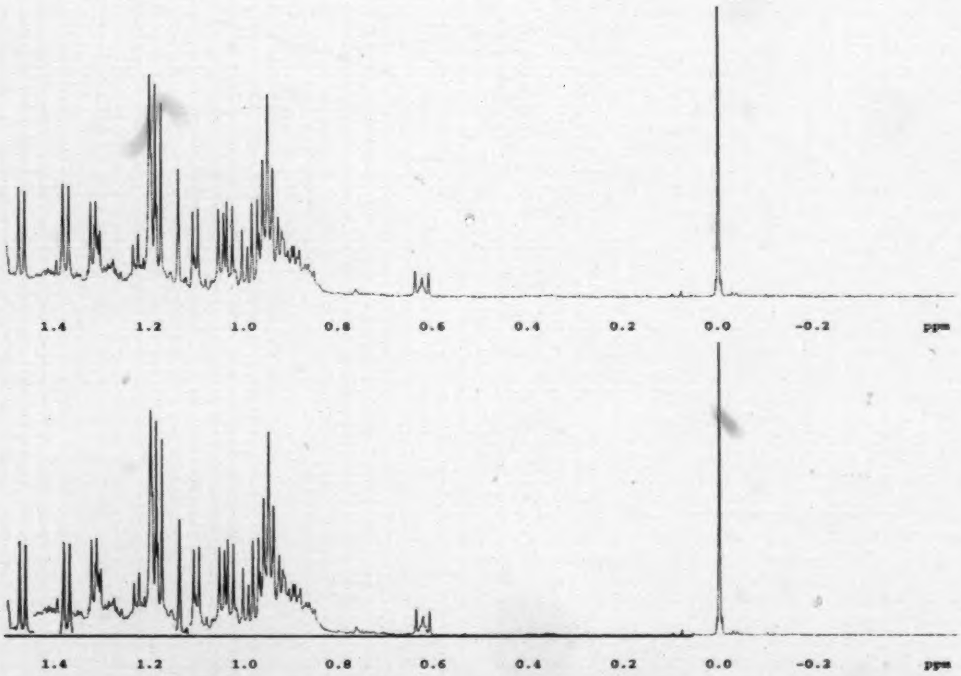


Figure C-7. **Quantification by circumventing baseline.** Graphical representation of the conceptual removal of a rolling or broadened baseline from selected metabolite peaks before concentration determination by resonance integration.

Through a collaboration with Dr. Shah in the Faculty of Engineering, UofA, an automated processing program was created to model and remove excess baseline features while retaining the resonant peaks to be integrated. Further research is continuing to optimize such processes and determine the best solution to deal with baseline issues and improve quantitative accuracy.

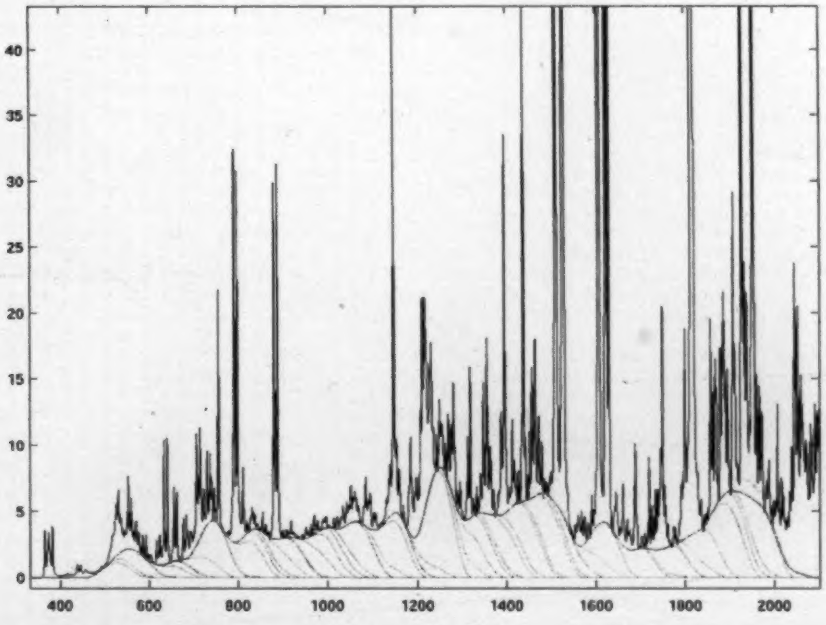


Figure C-8. **Post-acquisition baseline removal by computer modeling.** Graphical demonstration of baseline fitting and removal from 1D ^{1}H-NMR spectrum of urine by software created through a collaboration with Dr. Shah. Baseline fitting is optimized to remove the broad-rolling features while retaining the sharper metabolite peaks.

C.6 Quantitation Macro 'Lines'

Due to the limitation of resonance integration for quantitative calculations in situations of peak overlap and baseline perturbations a macro was written that allows peaks to be automatically picked, integration windows set (without peak overlap), and peak quantitation calculated with reduced influence of the baseline.

The macro was called Lines. Work is continuing to define its efficacy for quantitative accuracy. Lines works by picking peaks above a certain threshold specified by the user, automatically places integration windows of a specified width as determined by the linewidth, and integrates that region.

```
"Lines Program"
"$number is the number of peaks found in the spectrum"
"$height is the height of respective peak"
"$position is the position (in HZ) of respective peak"
"user added variable for control of linewidth at a particular peak height e.g. 0.1 width at 10% of height"
clear "clears bottom display"
if ($#>0) then "determines if the user has entered a value, if yes $#=1, if no $#=0"
$measure=$1 "this assigns $measure to the value entered by the user"
else
write('alpha','Default of 0.5 height used for linewidth determination')
$measure=0.5
endif
"line below for proper integration values"
fn=132k lb=0.5 ins=1 io=50 insref=1.3e-6 is=1e5 vs=10000 wft
$i=1"variable used to control loop"
write('reset',userdir+'/../line.txt') "makes sure file isn't already there"
nll:$number "nll vnmr command finds frequencies and heights"
repeat
"if it's the first line setup the header"
if $i = 1 then
getll(1):$height,$position
dres($position,$measure):$linewidth,$resolution
mark($position+($linewidth*0.5),$linewidth):$markht,$int
write('alpha','#    Height   Position    Linewidth  Resolution Integral')
write('alpha','%4.0f%10.4f%10.4f%10.4f%10.4f%10.6f',$i,$height,$position,$linewidth,$resolution,$int)
write('file',userdir+'/../line.txt','#    Height   Position    Linewidth  Resolution Integral')
write('file',userdir+'/../line.txt','%4.0f%10.4f%10.4f%10.4f%10.4f%10.6f',$i,$height,$position,$linewidth,$resolution,$int)
else
getll($i):$height,$position  "specifically gets one of the peaks"
dres($position,$measure):$linewidth,$resolution
mark($position+($linewidth*0.5),$linewidth):$markht,$int
write('alpha','%4.0f%10.4f%10.4f%10.4f%10.4f%10.6f',$i,$height,$position,$linewidth,$resolution,$int)
write('file',userdir+'/../line.txt','%4.0f%10.4f%10.4f%10.4f%10.4f%10.6f',$i,$height,$position,$linewidth,$resolution,$int)
endif
$i=$i+1
until $i>$number
```

Figure C-9. **Lines macro created for automated quantitation**. The macro 'Lines' automatically selects metabolite resonance peaks, determines peak height and linewidth, sets integration regions to a user-specified linewidth, and automatically returns a table outlining every peak and the calculated integration.

C.7 Biochemical Definition of Baseline

Various biochemical techniques were used to isolate, identify, and determine a 'average normal' urine baseline component as visualized in 1D ^{1}H-NMR spectrum. Urine was sampled from normal subjects and different biochemical components of the urine were isolated. Urine was spun, filtered, lyophilized, dialyzed, and underwent a Folch extract. Emulsified tissue samples underwent trichloroacetic acid extraction. The different separation and isolation techniques allowed for a general understanding of lipid, protein, salt, and metabolite content in normal urine. This aided in defining the source of NMR baseline rolls and will assist in the future development of methods to remove the baseline effects from metabolite quantitation.

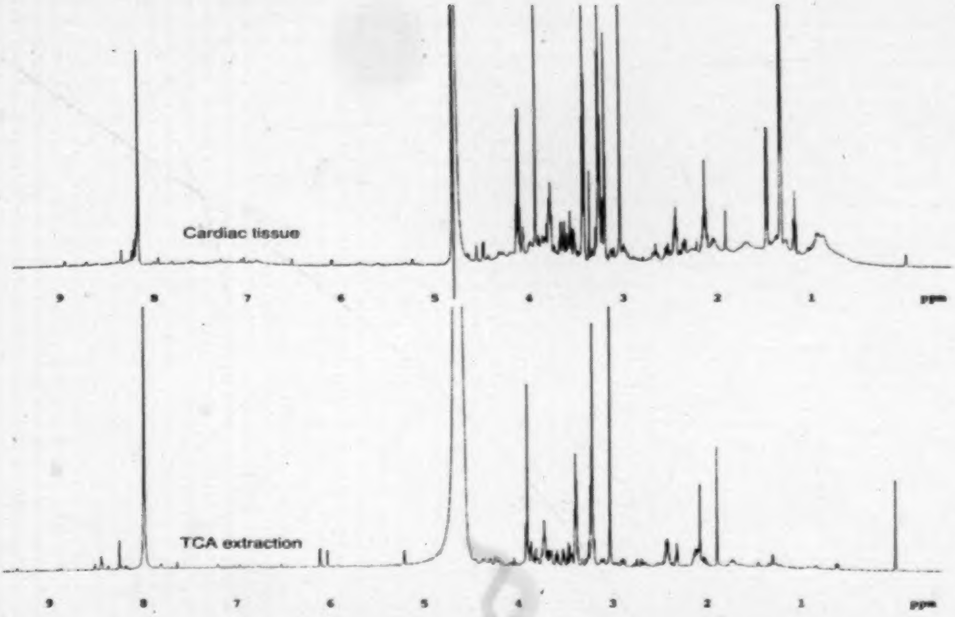


Figure C-10. **Pre-acquisition chemical removal of NMR baseline**. 1D ^{1}H-NMR spectra of emulsified rabbit cardiac tissue before and after TCA extraction. Baseline is improved following extraction, but with this method some metabolite peaks are also lost.

C.8 MS of Urine Macromolecules

The normal human urine samples were submitted for HPLC and mass spectrometry analysis in order to identify some of the major macromolecular components of urine. A few automated 'hits' were confirmed following a MASCOT database search, however, many of the peaks remain unidentified. Work is continuing, largely in collaboration with the Human Metabolome Database, to broaden the number of identified protein and lipid components of normal human urine.

CONCLUSION

Baseline is a significant challenge with NMR qualitative and quantitative analysis. The baseline hides low-concentration molecules and distorts quantitative accuracy. Methods to remove the effects of spectral baseline include sample extraction of large molecules, pulse sequence removal of large molecule resonances, and post-acquisition processing such as computer modeling of baselines and different quantitation software that circumvents the influence of the baseline. My research is continuing to identify and refine the best method for removing the influence of a 'rolling' baseline. In the meantime researchers should be aware of these considerations so that errors in quantitation are kept to a minimum.

APPENDIX D

Binding kinetics of calcium, cardiac troponin I peptide, and cardiac troponin C

OVERVIEW

This appendix reviews the first cardiac protein structure and function project I participated in upon entering the Sykes lab January of 2001. Ca^{2+} and human cardiac troponin I (cTnI) peptide binding to human cardiac troponin C (cTnC) have been investigated with the use of 2D $\{^1H, {}^{15}N\}$-HSQC NMR spectroscopy. The spectral intensity, chemical shift, and line-shape changes were analyzed to obtain the dissociation (K_D) and off rate (k_{off}) constants at 30°C. The results show that sites III and IV exhibit 100 fold higher Ca^{2+} affinity than site II ($K_{D(III,IV)} = \sim 0.2$ μM, $K_{D(II)} = \sim 20$ μM), but site II is partially occupied before sites III and IV are saturated. The addition of the first two equivalents of Ca^{2+} saturates 90% of sites III and IV and 20% of site II. This suggests that the Ca^{2+} occupancy of all three sites may contribute to the Ca^{2+}-dependent regulation in muscle contraction. We have determined a k_{off} of 5000 s^{-1} for site II Ca^{2+} dissociation at 30°C. Such a rapid off rate had not yet been measured. Three cTnI peptides, $cTnI_{34-71}$, $cTnI_{128-147}$, and $cTnI_{147-163}$ were titrated to Ca^{2+}-saturated cTnC. In each case, the binding occurs with a 1:1 stoichiometry. The determined K_Ds and k_{off}s are ≤ 1 μM and 5 s^{-1} for $cTnI_{34-71}$, 78 ± 10 μM and 5000 s^{-1} for $cTnI_{128-147}$, and 150 ± 10 μM and 5000 s^{-1} for $cTnI_{147-163}$, respectively. Thus, the dissociation of Ca^{2+} from site II and $cTnI_{128-147}$ and $cTnI_{147-163}$ from cTnC are rapid enough to be involved in the contraction-relaxation cycle of cardiac muscle, while that of $cTnI_{34-71}$ from cTnC may be too slow for this process.

INTRODUCTION

As the Ca^{2+} binding member of the troponin complex, troponin C (TnC) plays a key role in the Ca^{2+} regulation of contraction/relaxation in striated muscle. Conformational changes in TnC induced by Ca^{2+} association/dissociation are believed to be transmitted through other thin filament proteins, troponin I, troponin T, tropomyosin, and actin, resulting in activation/inhibition of actomyosin ATPase and muscle contraction/relaxation (see [1] [2]for reviews). Two isoforms of TnC exist in striated muscle, skeletal muscle troponin C (sTnC) and cardiac muscle TnC (cTnC). Both molecules are dumbbell-shaped (see [3] for a review) with two domains, N- and C-, connected through a linker and comprise four EF-hand helix-loop-helix motifs as potential Ca^{2+}-binding sites (sites I-IV). Sites I and II are paired as a unit in the N-terminal half, and sites III and IV form another pair in the C-terminal half of the molecule. Sites III and IV are of relatively higher affinity for Ca^{2+} and also bind Mg^{2+}. Sites I and II are of lower affinity and are believed to be specific for Ca^{2+}. Site I in cTnC is unable to bind Ca^{2+} at physiological concentrations due to key amino acid substitutions [4]. Current evidence indicates a largely structural role for the C-domain whose sites would be occupied by Ca^{2+}/Mg^{2+} throughout the contraction/relaxation cycle. The regulatory role is considered to be associated with the conformational changes induced by the association and dissociation of Ca^{2+} from N-domain sites I and II of sTnC or site II of cTnC.

The fundamental difference between sTnC and cTnC is that the binding of Ca^{2+} to sTnC is coupled by a large structural 'opening'[5], while the association of Ca^{2+} with cTnC results in minimal conformational changes[6, 7]. This is mainly due to the fact that both sites I and II are functional in sTnC [8], while only site II is active in cTnC [9]. The exposed hydrophobic surface in the Ca^{2+}-saturated sTnC has long been proposed [10] and has subsequently been proven [11, 12] as the sTnI binding site. The significant reduction in the hydrophobic surface exposure of Ca^{2+}-saturated cTnC suggested that the mode of interaction between cTnC-cTnI may be different than that between sTnC-sTnI. However, we have found that both the regulatory domains of sTnC and cTnC adopt similar 'open' conformations when bound to their respective TnI regions ($sTnI_{115-131}$ and $cTnI_{147-163}$)[13, 14]. This region of TnI has been identified by many biological and biophysical studies to be the region responsible for binding to the regulatory domain of TnC and this interaction modulates the interaction between the N-terminal ($sTnI_{1-40}$ and $cTnI_{34-71}$) and the inhibitory regions ($sTnI_{96-115}$ and $cTnI_{128-147}$) of TnI and TnC (see [15, 16] for a review). Thus, the sequence of events involved in initiating skeletal and cardiac muscle contraction are actually very similar. However, the kinetics and thermodynamics of these events must differ for the two systems to account for the different physiological behavior of the two muscle types (see [17, 18]for discussions).

To understand the unique delicate energetic balance that exists for each system, it is important to study the time scales of Ca^{2+} binding and release from TnC, the accompanied structural changes, and the subsequent interactions of TnI with TnC. A number of studies have been performed on the kinetics of Ca^{2+} binding to TnC, isolated or in the troponin complex. There are major differences between the skeletal and cardiac isoforms. In sTnC, Ca^{2+} binding to sites I/II appears to be diffusion limited (k_{on} = ~ 10^8 $M^{-1}s^{-1}$) with a Ca^{2+} dissociation rate (k_{off}) of ~ 400-500 s^{-1} and the conformational change occurs almost simultaneously with the Ca^{2+} association/dissociation[19-21]. In cTnC, although Ca^{2+} binding to site II is also diffusion limited (k_{on} = ~ 10^8 $M^{-1}s^{-1}$, k_{off} = ~ 500-800 s^{-1}), the conformational change was found to be significantly slower than the Ca^{2+} on- and off-rates[22-24]. These reported on- and off- rates may not, however, represent the time scales of the molecular events that occur at physiological temperature (*e.g.* 37°C in human heart), since most of these experiments were performed at 4°C because Ca^{2+} kinetics become too fast to be measured by stopped flow methods at temperatures higher than 4°C [24].

In this appendix, I examined the kinetics of Ca^{2+} and three cTnI peptides ($cTnI_{34-71}$, $cTnI_{128-147}$, and $cTnI_{147-163}$) binding to cTnC at 30°C with the use of 2D {^1H, ^{15}N} HSQC NMR spectroscopy. An advantage of NMR spectroscopy is that this technique can measure the time scales of the molecular events on proteins at temperatures close to physiological and can report information related to individual atoms throughout the sequence. Ca^{2+} and peptide titrations of sTnC and cTnC domains followed in detail by 2D {^{1}H, ^{15}N} HSQC NMR spectroscopy have been proven to be a powerful way of providing information such as binding stoichiometry, affinity, and energetics. For example, using this technique, we have previously characterized Ca^{2+} binding to sNTnC and sCTnC[25, 26], to the cNTnC and an E41A mutant of sNTnC [27]. We also monitored the binding of $sTnI_{115-131}$ to sNTnC and E41A sNTnC[11, 17], $cTnI_{147-163}$ to cNTnC [13], $sTnI_{1-40}$ and $sTnI_{96-115}$ to sCTnC [26], and $cTnI_{34-71}$ and $cTnI_{128-147}$ to cCTnC [28]. In this study, the analysis of NMR spectral intensity, chemical shift, and line-shape changes induced by Ca^{2+} binding to cTnC and cTnI peptide binding to Ca^{2+}-saturated cTnC allow us to determine the dissociation (K_D) and off rate (k_{off}) constants at 30°C for these binding events. These measurements provide new insights into the time course of the important molecular events involved in the regulation of cardiac muscle contraction. In turn, it allows us to understand further the differences between cardiac and skeletal muscle contraction.

EXPERIMENTAL PROCEDURES

D-A. Sample Preparation

Recombinant human cTnC (residues 1-161) with the mutations C35S and C84S [denoted cTnC (C35S,C84S)] was used in this study. Since cTnC (C35S,C84S) is the only protein used throughout this work, (C35S,C84S) is omitted in this paper. The engineering of the expression vector of cTnC (C35S, C84S) was as follows: pET3a.cTnC(WT) was constructed as described previously [18] and is used as template for the construction of pET3a.cTnC(C35S,C84S); two rounds of mutagenesis (Stratagene) were performed using paired 31-mer oligonucleotides 5'-GGCGCTGAGGATGGCAGCATCAGCACCAAGG-3' and 5'-CCTTGGTGCTGATGCTGCCATCCTCAGCGCC-3' for C35S mutation and paired 37-mer primers 5'-CCTGGTCATGATGGTTCGCAGCATGAAGGACGACAGC-3' and 5'-GCTGTCGTCCTTCATGCTGCGAACCATCATGACCAGG-3' for C84S mutation (with base changes underlined). DNA sequence analysis was used to confirm the correctness of all the mutants. The expression and purification of the ^{15}N-cTnC in BL21(DE3)pLysS cells were as described previously for ^{15}N-cNTnC [27]. Decalcification of ^{15}N-cTnC follows the procedure for ^{15}N-cNTnC [27] except that 200 mM EDTA was used instead of 100 mM. A higher concentration of EDTA helps to remove the tightly bound metal ions in the C-domain of cTnC. Three synthetic cTnI peptides, cTnI$_{34-71}$, Acetyl-AKKKSKISASRKLQLKTLLLQIAKQELEREAEERRGEK-Amide, cTnI$_{128-147}$, Acetyl-TQKIFDLRGKFKRPTLRRVR-Amide, and cTnI$_{147-163}$, Acetyl-RISADAMMQALLGARAK-Amide, respectively, were prepared using standard methodology for a typical TnI peptide [29]. The sequences were confirmed by amino acid analysis [30] and the mass verified by electrospray mass spectrometry. All NMR samples were 500 μL in volume. The buffer conditions were 100 mM KCl, 10 mM imidazole, 0.2 mM 2,2-dimethyl-2-silapentanesulfonic acid (DSS), and 0.01% NaN_3 in 90% H_2O/10% D_2O, and the pH was 6.9. The concentration of the apo ^{15}N-cTnC sample used for Ca^{2+} titration was determined to be 0.95 mM. The concentration of the ^{15}N-cTnC sample used for cTnI$_{34-71}$ titration followed by cTnI$_{147-163}$ were 0.80 mM and the ^{15}N-cTnC samples used for cTnI$_{128-147}$ titrations were 0.56 mM. Each of the sample contains ~ 3-5 mM $CaCl_2$. Gilson Pipetman P (model P2) was used to deliver the $CaCl_2$ solution and cTnI$_{34-71}$ and cTnI$_{128-147}$ peptide solutions for titrations. Solid cTnI$_{147-163}$ was added during titration.

D-B. Ca^{2+} titration of ^{15}N-cTnC

Stock solutions of standardized 50 mM and 100 mM $CaCl_2$ in water were used for the titrations. To a NMR tube containing a 500 μL sample of 0.95 mM ^{15}N-cTnC were added consecutively aliquots of 0.5, 1.0, and 1.5 μL of 50 mM $CaCl_2$ for the first three titration points, and aliquots of 1.0, 1.0, 1.0, 1.0, 1.5, 1.5, 1.5, 1.0, 1.0, 1.0, 1.0, 2.0, 2.0, 3.0 μL of 100 mM $CaCl_2$ were added consecutively for the next 14 individual titration points. The sample was mixed thoroughly with each addition (total of 17 additions). The total volume increase was 22.5 μL, and the change in protein concentration due to dilution was taken into account for data analysis. The change in pH from Ca^{2+} addition was negligible. Both 1D ^{1}H and 2D {^{1}H, ^{15}N}-HSQC NMR spectra were acquired at every titration point.

D-C. cTnI$_{34-71}$ titration of ^{15}N-cTnC•3Ca^{2+}

To a NMR tube containing a 500 μL sample of 0.80 mM ^{15}N-cTnC and 5 mM $CaCl_2$ were added consecutive aliquots of 1.0 μL of 31 mM cTnI$_{34-71}$ stock solution in water for the 13 individual titration points. The sample was mixed thoroughly with each addition. The total volume increase was 13 μL, and the change in protein concentration due to dilution was taken into account for data analysis. The decrease in pH associated with $cTnI_{34-71}$ additions were adjusted by 1 M NaOH to pH 6.9 at every titration point. Both 1D ^1H and 2D $\{^1H, ^{15}N\}$-HSQC NMR spectra were acquired at every titration point.

D-D. $cTnI_{147-163}$ titration of ^{15}N-$cTnC{\bullet}3Ca^{2+}{\bullet}cTnI_{34-71}$

Following titration B, the sample was filtered, and the appropriate amount of NMR buffer was added to generate a 500 μL NMR sample containing the $cTnC{\bullet}3Ca^{2+}{\bullet}cTnI_{34-71}$ complex. The total cTnC concentration in this sample is 0.75 mM. $cTnI_{147-163}$ peptide is highly soluble in aqueous solution but tends to form a gel at high concentrations, likely due to aggregation. Thus, no stock peptide solution was prepared; instead, solid peptide was added at every titration point (total 6 additions). The concentrations of ^{15}N-cTnC and $cTnI_{147-163}$ were determined by amino acid analysis [30] at every titration point, giving the peptide/protein ratios. The decrease in pH associated with $cTnI_{147-163}$ additions was adjusted by 1 M NaOH to pH 6.9. Both 1D ^1H and 2D $\{^1H, ^{15}N\}$-HSQC NMR spectra were acquired at every titration point.

D-E. $cTnI_{128-147}$ titration of ^{15}N-$cTnC{\bullet}3Ca^{2+}$

This titration was as described previously [31]. In the previous report, the binding affinity (K_D) was determined but the line shape analysis was not done.

The titration data were analyzed further here to determine the exchange broadening and off rate constant (k_{off}), for the purpose of comparing with the results of $cTnI_{34-71}$ and $cTnI_{147-163}$ binding to cTnC.

D-F. NMR Spectroscopy

All of the NMR spectra were obtained at 30°C using Unity 600 MHz and Unity Inova 500 MHz spectrometers. 2D $\{^1H, {}^{15}N\}$-HSQC NMR spectra were acquired using the sensitivity-enhanced gradient pulse scheme developed by Lewis E. Kay and co-workers[32, 33]. The ^1H and ^{15}N sweep widths were 7000 and 1500 Hz, respectively on the 500 MHz spectrometer and were 8000 and 1650 Hz, respectively on the 600 MHz spectrometer. All spectra were processed and analyzed using VNMR (Varian Associates) and NMRPipe [34] and referenced according to the IUPAC conventions.

RESULTS

D-A. Ca^{2+} titration of cTnC

This study involves the use of 2D $\{^1H,{}^{15}N\}$ HSQC NMR spectroscopy to characterize Ca^{2+} and TnI peptide binding to the full length cTnC. In previous studies, we have demonstrated the utility of 2D $\{^1H,{}^{15}N\}$ HSQC NMR spectroscopy in characterizing Ca^{2+} and TnI peptide binding to the isolated N- or C- domains of sTnC or cTnC (see [17]and references therein). The 2D $\{^1H, {}^{15}N\}$-HSQC NMR spectrum of Ca^{2+}-saturated cTnC is completely assigned [6] and used as a guide to assign the Ca^{2+}-induced spectral changes. Figure F-1 depicts the Ca^{2+}-induced 2D $\{^1H, ^{15}N\}$-HSQC NMR spectral changes of backbone amide resonance in cTnC. The spectrum of apo cTnC is shown as multiple contours in Figure D-1A, in which the N-domain peaks (*e.g.* G70 and G42) are well dispersed, while the C-domain peaks fall within regions of ^1H and ^{15}N chemical shifts characterized as "random coil" by Wuthrich [35]. This suggests that the apo C-domain does not adopt a defined structure in solution, while the apo N-domain possesses a folded structure. Sites III/IV possess higher affinity than site II, so the C-domain is filled first by Ca^{2+} and this binding occurs with slow exchange kinetics on the NMR time scale. Thus, as titration progresses, the resonance peaks corresponding to a structured C-domain appear and grow (single contour peaks in Figure D-1A).

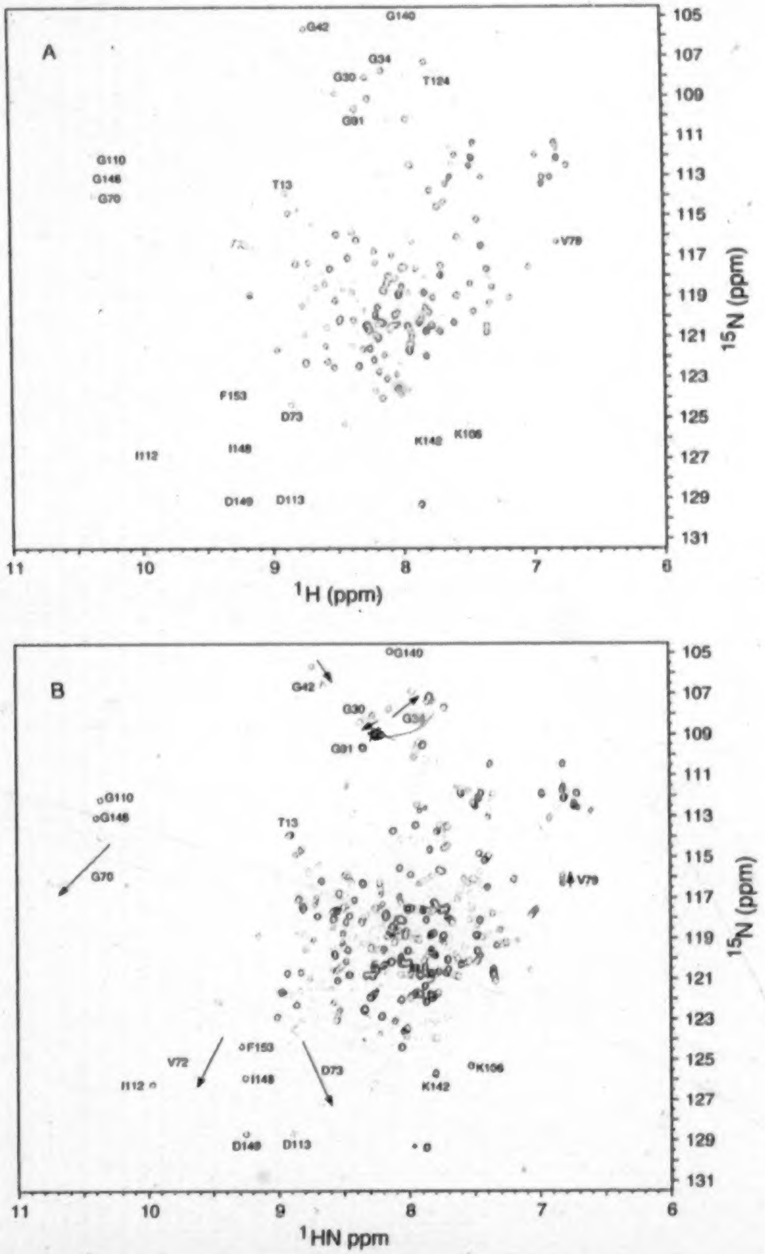


Figure D-1. **Titration of human cTnC with Ca^{2+}.** Titration of human intact cTnC with Ca^{2+} as monitored by 2D {^{1}H, ^{15}N}-HSQC NMR spectra of the backbone amide regions of cTnC. (A) The cross peaks corresponding to apo cTnC are shown as multiple contours, whereas the peaks corresponding to the $[\mathrm{Ca}^{2+}]_{\mathrm{total}}/[\mathrm{cTnC}]_{\mathrm{total}}$ ratio of 1.47 are shown as single contours. (B) The peaks corresponding to the $[\mathrm{Ca}^{2+}]_{\mathrm{total}}/[\mathrm{cTnC}]_{\mathrm{total}}$ ratio of 1.47 are shown in multiple contours and the spectra representing various Ca^{2+} additions after the $[\mathrm{Ca}^{2+}]_{\mathrm{total}}/[\mathrm{cTnC}]_{\mathrm{total}}$ ratio of 1.47 are superimposed and shown as single contours.

This is better illustrated by the growth of D113 and D149 peaks in Figure D-2. It is interesting to notice that this growth did not stop when the $[Ca^{2+}]_{total}$:$[cTnC]_{total}$ ratio reached 2:1 and the resonance peaks corresponding to the N-domain started shifting at $[Ca^{2+}]_{total}$:$[cTnC]_{total}$ ratio of 1.47:1 (Figure D-1B). G70 and V72 are typical examples (Figure D-2). This indicates that site II is being partially filled before sites III/IV are completely saturated. When the intensities of D113 and D149, respectively, and the chemical shifts (both ^1H and ^{15}N) of G70 and V72, respectively, were measured at every titration point and the changes were averaged, normalized, and plotted as a function of $[Ca^{2+}]_{toal}/[cTnC]_{total}$ ratio, curve A (Figure D-3) corresponding to Ca^{2+} binding to sites III/IV increases rapidly and in a parallel fashion during the addition of the first and second equivalent of Ca^{2+}. During the addition of the third equivalent of Ca^{2+}, it increases slowly to reach 100% occupancy. On the other hand, the gradient of curve B (Figure D-3) corresponding to Ca^{2+} binding to site II increases slowly during addition of the first 2 equivalents of Ca^{2+} and increase rapidly in the region between 2 and 3 equivalent to reach 100% occupancy. The curves indicate that all three sites have some degree of occupancy from the first addition of Ca^{2+} and that the occupancy of all three sites is essentially completed when 3 equivalents of Ca^{2+} have been added.

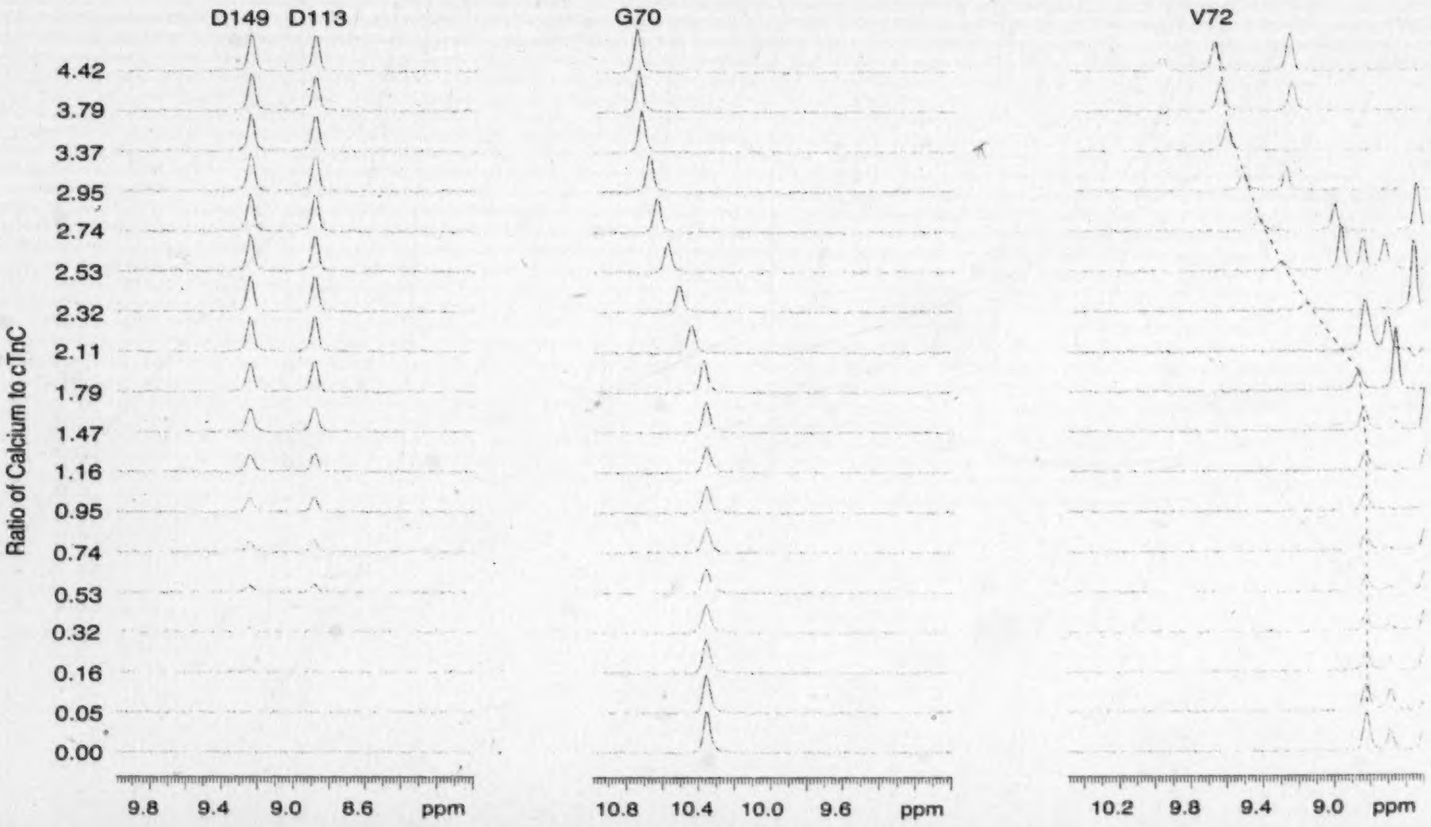


Figure D-2. **One-dimensional traces demonstrating Ca^{2+} binding.** Stacked plots of 1D (^1H dimension) traces from 2D {^1H, ^{15}N}-HSQC NMR spectral cross peaks of cTnC residues D149, D113, G70, and V72. D149 and D113 represents Ca^{2+} binding to the C-domain, while G70 and V72 represents Ca^{2+} binding to the N-domain.

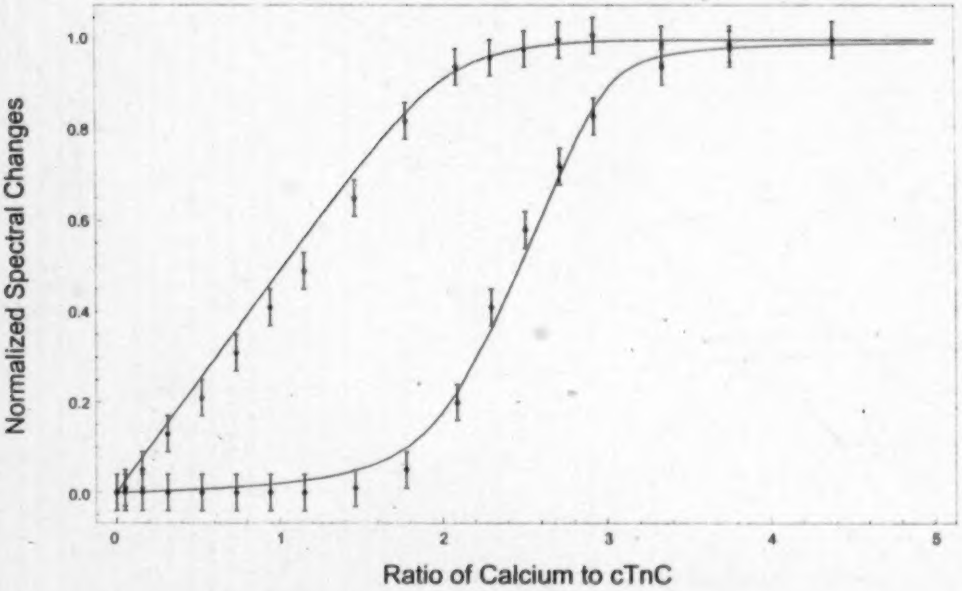


Figure D-3. **Calcium binding curves for the C- and N-domains of cTnC.** Binding curves derived from the 2D {^{1}H, ^{15}N}-HSQC NMR spectra of cTnC upon addition of Ca^{2+}. The A curve is normalized according to the average intensity changes of D149 and D113, representing Ca^{2+} binding to the C-domain, while the B curve is normalized according to averaged chemical shift changes of G70 and V72, representing Ca^{2+} binding to the N-domain. The best-fit curves to the data are shown by solid lines. The curve fitting procedures are as described in Results.

Assuming sites III and IV in the C-domain are independent and equal, which is not distinguished from 100% positive cooperativity and site II is independent of sites III and IV, and assuming each of the three sites (II, III, IV) binds Ca^{2+} in a 1:1 stoichiometry, the spectral changes as a function of $[Ca^{2+}]_{total}/[cTnC]_{total}$ were fit to the equations

$$\text{Site III} + Ca^{2+} \leftrightarrow \text{Site III}{\bullet}Ca^{2+}$$
$$\text{Site IV} + Ca^{2+} \leftrightarrow \text{Site IV}{\bullet}Ca^{2+}$$
$$\text{Site II} + Ca^{2+} \leftrightarrow \text{Site II}{\bullet}Ca^{2+}$$

and yielded macroscopic dissociation constants of $K_{D(III,IV)}$ = ~ 0.2 μM for sites III and IV and $K_{D(II)}$ = ~ 20 μM for site II, respectively. This fitting also permits the calculation of a site III and IV occupancy of ~ 90% and a site II occupancy of ~ 20% at 2 equivalents of Ca^{2+} (The curve fitting script used is available upon request from brian.sykes@ualberta.ca). These results are consistent with the consensus results that the C-domain sites III/IV have ~100 fold higher Ca^{2+} binding affinity than the N-domain site II. However, the determined site II dissociation constant of K_D = ~ 20 μM is approximately 10 times weaker than that for native cTnC reported previously[24, 36-38]. This discrepancy is probably due to the fact that the previous reported affinities were measured at lower temperatures (*e.g.* 4°C) and it is likely that site II binds Ca^{2+} tighter at lower temperatures than 30°C. Ca^{2+} titration of cTnC at lower temperatures (*e.g.* 4°C) by the use of 2D {^1H,^{15}N} HSQC NMR spectroscopy would provide insights into this matter, a project currently in progress in our laboratory. Somewhat surprisingly, this affinity is also ~10 times weaker than what we have observed in the Ca^{2+} titration of cNTnC (K_D = 2.6 μM) [27]. It seems that the presence of the C-terminal domain somehow reduced the affinity of Ca^{2+} for the N-domain of cTnC. At present, there is no compelling data for rationalization.

In addition to the Ca^{2+}-induced intensity and chemical shift changes, individual cross peaks in the 2D {^1H, ^{15}N} HSQC NMR spectra during Ca^{2+} titration display differential broadening, especially those in the N-domain, indicating exchange broadening. Typical examples are G70 and V72 as shown in Figure D-2. The NMR spectral changes can occur in fast, intermediate, and/or slow exchange limit on the NMR time scale depending on the size of the Ca^{2+}-induced resonance shift [25]. In the intermediate to fast exchange limit, the effect of chemical exchange on line width is dependent on the chemical shift differences between free and the bound species:

$$\Delta v_{ex} = P_f P_b \tau_{ex} (\Delta\delta_i)^2$$

where Δv_{ex} is the observed line width, P_f and P_b are the populations of free and bound cTnC, and τ_{ex} is the exchange lifetime defined as $(\tau_f \tau_b)/(\tau_f + \tau_b)$. Previously, we have used the line width simulation to analyze the behavior of the 2D $\{^1H,^{15}N\}$ HSQC NMR spectra taken during titrations of sTnC with sTnI peptides[11, 12]. Here we use the same approach to analyze the 2D $\{^1H,^{15}N\}$ HSQC NMR spectral changes induced by Ca^{2+} binding to the N-domain of cTnC. One dimensional 1D traces in the ^1H-dimension through the cross peaks of residues G42, D73, G70, and V72 are shown in Figure D-4 to demonstrate the effect of exchange at various titration points during the addition of Ca^{2+}. For G42, where the $\Delta\delta$ = 43 Hz, the line shape only slightly broadens during the titration. On the other hand, V72, which has an $\Delta\delta$ of 424 Hz, broadens substantially during the titration and then sharpens dramatically at the end. The $\Delta\delta$s for D73 and G70 are 92 Hz and 201 Hz, respectively, and the line broadening fell between G42 and V72. The line shape changes for these four residues were simulated to obtain the site II Ca^{2+} off rate constant k_{off}. The line widths of the free ($\Delta\nu_f$) and the bound ($\Delta\nu_b$) species and the starting (δ_f) and the final (δ_b) resonance positions for those four residues are listed in the legend of Figure D-4. The simulation were done using the experimentally derived values for $\Delta\nu_f$, $\Delta\nu_b$, δ_f, δ_b, and $K_{D(II)} = 20$ μM and $K_{D(III/IV)} = 0.2$ μM and adjusting k_{off}. A k_{off} of 5000 s^{-1} provides the closest fit to the experimental data for all four residues (Figure D-4). Since the simulation program does not take into account the relaxation in both ^1H and ^{15}N dimensions during the 2D $\{^1H, ^{15}N\}$ HSQC pulse sequence, which will lead to lower intensities in the middle of the titration when the line broadening is the largest, differential intensities of the cross peaks between the simulated and experimental spectra are observed (Xu Wang, Monica X. Li, and Brian D. Sykes, unpublished data). The most obvious is with V72 (Figure D-4). The site II Ca^{2+} off rate of 5000 s^{-1} at 30°C had not been measured because this rate is too fast to be measured by fluorescence stopped flow experiments. Using the relationship, $K_D = k_{off}/k_{on}$, the calculated k_{on} is 2.5×10^8 $M^{-1}s^{-1}$, indicating that Ca^{2+} binding to site II is diffusion controlled at 30°C.

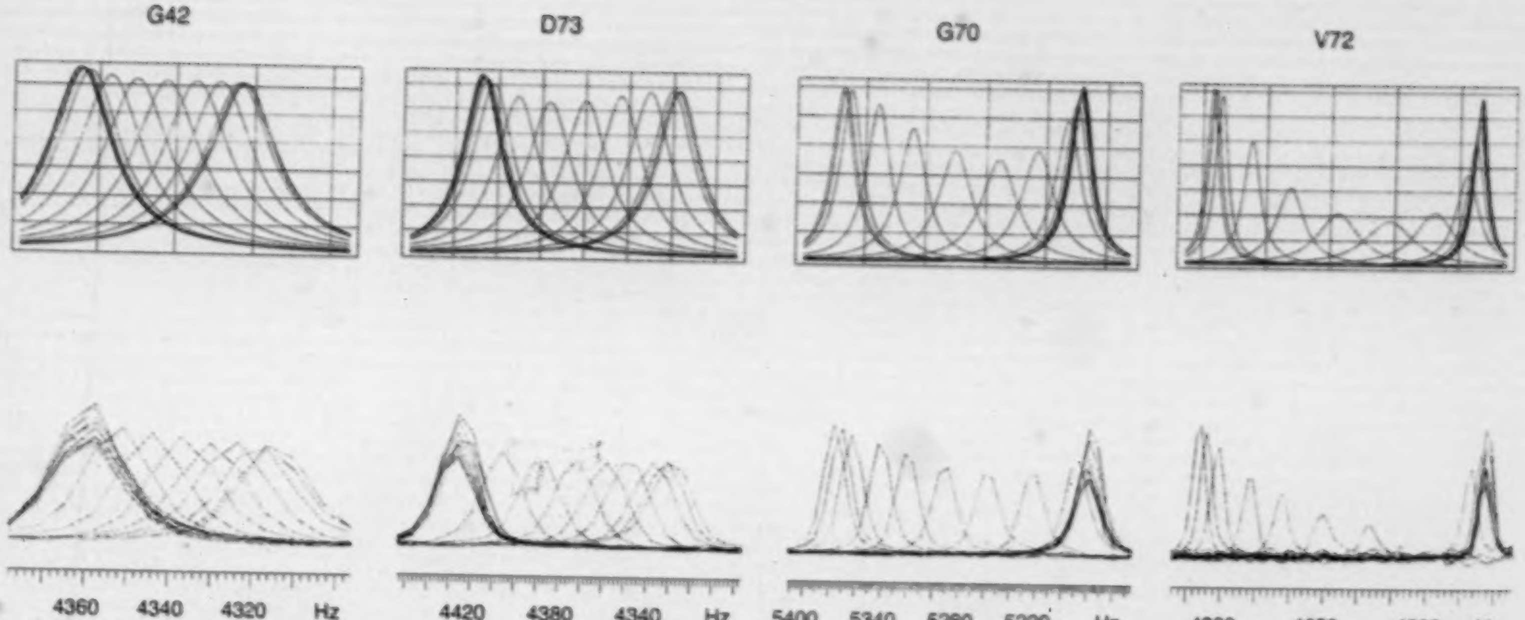


Figure D-4. **Computer simulated and empirical binding traces for Ca^{2+} and cTnC**. 1D (^{1}H dimension) traces taken through the 2D {^{1}H, ^{15}N}-HSQC NMR spectral cross peaks of G42, D73, G70, and V72 during Ca^{2+} titration are shown in lower panel. While the peaks of G42 and D73 shift from left to right, those of G70 and V72 shift from right to left. Upper panel are computer simulations using $K_{D(III/IV)} = 0.2$ μM, $K_{D(II)} = 20$ μM, and $k_{off} = 5000\ s^{-1}$. The line widths of the free (Δv_f) and the bound (Δv_b) species and the starting (δ_f) and the final (δ_b) resonance positions for those four residues are as following: G42, $\Delta v_f = 19$ Hz, $\Delta v_b = 21$ Hz, $\delta_f = 4358$ Hz, and $\delta_b = 4315$ Hz; D73, $\Delta v_f = 21$ Hz, $\Delta v_b = 23$ Hz, $\delta_f = 4424$ Hz, and $\delta_b = 4332$ Hz; G70, $\Delta v_f = 21$ Hz, $\Delta v_b = 21$ Hz, $\delta_f = 5178$ Hz, and $\delta_b = 5379$ Hz; V72, $\Delta v_f = 22$ Hz, $\Delta v_b = 23$ Hz, $\delta_f = 4409$ Hz, and $\delta_b = 4833$ Hz.

D-B. cTnI$_{34\text{-}71}$ titration of cTnC•3Ca^{2+}

Previously, we have shown that cTnI$_{34\text{-}71}$ binds cCTnC•2Ca^{2+} in a 1:1 stoichiometry to form a stable cCTnC•2Ca^{2+}•cTnI$_{34\text{-}71}$ complex (K$_D$ $\leq$ 1 μM) and the interaction of cTnI$_{34\text{-}71}$ with cCTnC•2Ca^{2+} occurs with slow exchange kinetics on the NMR time scale [28]. Here we show a similar binding behavior of cTnI$_{34\text{-}71}$ for cTnC•3Ca^{2+}, indicating the interaction of this region of cTnI binds specifically to the C-domain of cTnC•3Ca^{2+} regardless of the presence of the N-domain. Figure D-5A shows a superimposition of the 2D {^{1}H, ^{15}N} HSQC NMR spectra of cTnC•3Ca^{2+} and the cTnC•3Ca^{2+}•cTnI$_{34\text{-}71}$ complex. As the titration progresses, the resonance peaks corresponding to cTnC•3Ca^{2+} becomes less intense while those corresponding to the cTnC•3Ca^{2+}•cTnI$_{34\text{-}71}$ complex grow. Only the C-domain residues are affected (Figure D-5A). For example, G110 and G146 shifted to new positions, while G70 peak did not move at all.

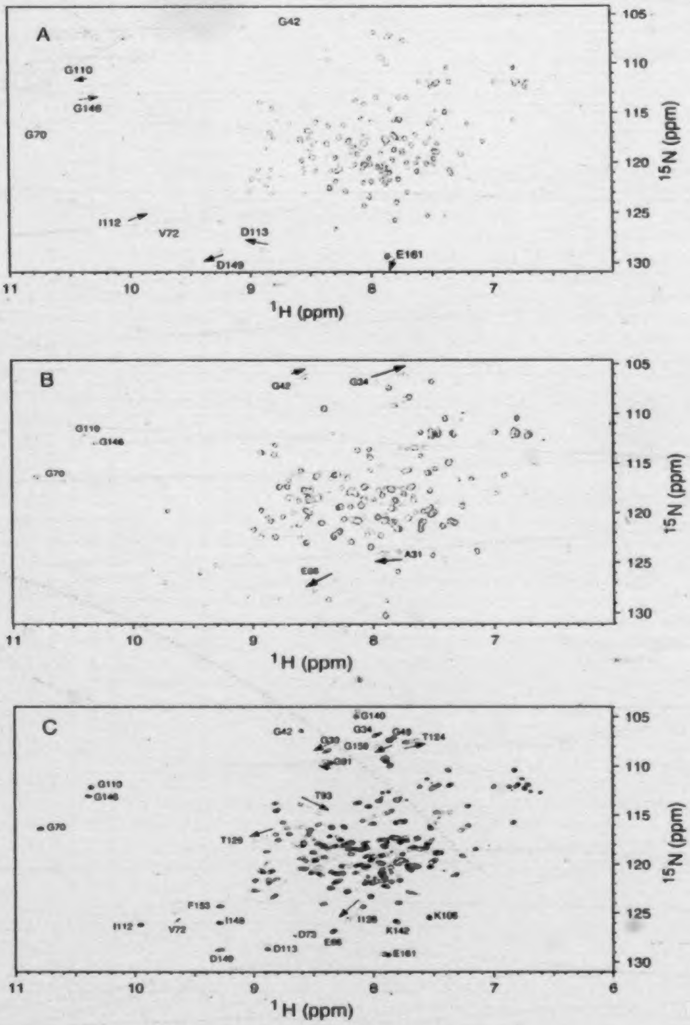


Figure D-5. **Titration of cTnC•3Ca^{2+} with cTnI peptides**. Titration of (A) cTnC•3Ca^{2+} with cTnI$_{34\text{-}71}$, (B) cTnC•3Ca^{2+}•cTnI$_{34\text{-}71}$ with cTnI$_{147\text{-}163}$, and (C) cTnC•3Ca^{2+} with cTnI$_{128\text{-}147}$ as monitored 2D {^{1}H, ^{15}N}-HSQC NMR spectra from the backbone amide regions of cTnC. In all three cases, the binding stoichiometry is 1:1. (A) The cross peaks corresponding to cTnC•3Ca^{2+} are shown as multiple contours, whereas those corresponding to cTnC•3Ca^{2+}•cTnI$_{34\text{-}71}$ are shown as single contours. (B) The cross peaks corresponding to cTnC•3Ca^{2+}•cTnI$_{34\text{-}71}$ are shown as multiple contours and the spectra representing various cTnI$_{147\text{-}163}$ additions are superimposed and shown as single contours. (C) The cross peaks corresponding to cTnC•3Ca^{2+} are shown as multiple contours and the spectra representing various cTnI$_{128\text{-}147}$ additions are superimposed and shown as single contours.

When the peptide to protein ratio reaches 1:1, all cross peaks corresponding to the C-domain of $cTnC{\bullet}3Ca^{2+}$ have completely disappeared while those corresponding to the complex attain maximum intensity. This phenomenon is more clearly illustrated in Figure D-6A. The intensity changes as a function of peptide to protein ratios were fit to the equation:

$$cTnC{\bullet}3Ca^{2+} + cTnI_{34\text{-}71} \leftrightarrow cTnC{\bullet}3Ca^{2+}{\bullet}cTnI_{34\text{-}71}$$

and yielded a dissociation constant (K_D) of ≤ 1 μM, agreeing with the affinity of $cTnI_{34\text{-}71}$ for $cCTnC{\bullet}2Ca^{2+}$ [28] and that of $sTnI_{1\text{-}40}$ for $sCTnC{\bullet}2Ca^{2+}$ [26]. The line shapes of G110 were simulated using experimentally derived values of $\Delta\nu_f$ = 20 Hz, $\Delta\nu_b$ = 24 Hz, δ_f = 6215 Hz, δ_b = 6329 Hz, $\Delta\delta$ = 114 Hz, and K_D = 1 μM and adjusting k_{off}. A k_{off} of 5 s^{-1} provides the closest fit to the experimental data (Figure D-7A). This indicates that the dissociation of this region of cTnI may be too slow to participate in muscle regulation.

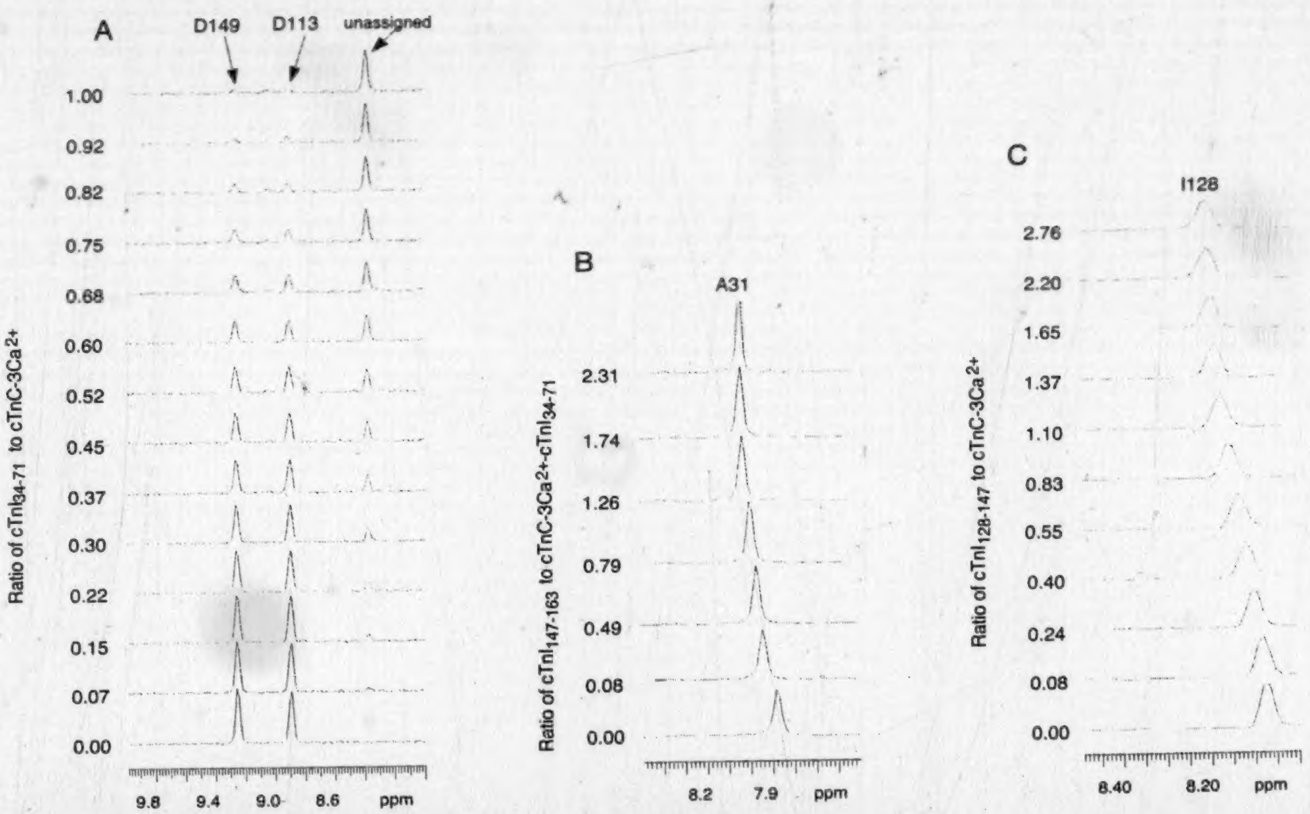


Figure D-6. **One-dimensional ^{1}H traces representing cTnI peptides binding to cTnC**. Stacked plots of 1D (^{1}H dimension) traces from 2D {^{1}H, ^{15}N}-HSQC NMR spectral cross peaks of cTnC residue G113 and D149, A31, and I128, respectively, representing (A) cTnI$_{34\text{-}71}$ binding to cTnC•3Ca^{2+}, (B) cTnI$_{147\text{-}163}$ binding to cTnC•3Ca^{2+}•cTnI$_{34\text{-}71}$, and (C) cTnI$_{128\text{-}147}$ binding to cTnC•3Ca^{2+}.

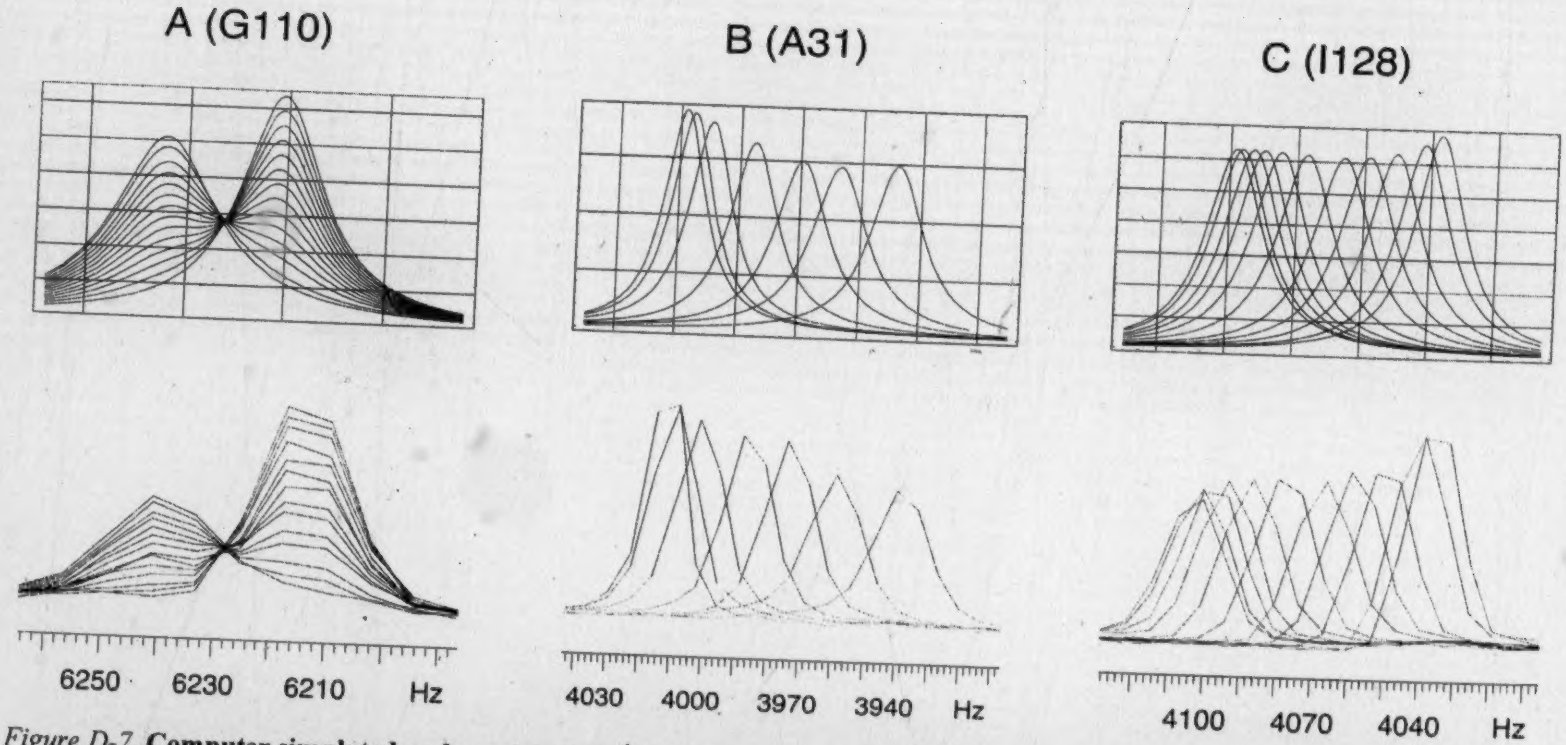


Figure D-7. **Computer simulated and empirical 1D ^{1}H-NMR traces for binding constants of cTnI peptides with cTnC**. 1D (^{1}H dimension) NMR spectral traces taken through the 2D {^{1}H, ^{15}N}-HSQC NMR spectral cross peaks of G110 in the titration of cTnC•3Ca^{2+} with cTnI$_{34-71}$ (A), A31 in the titration of cTnC•3Ca^{2+}•cTnI$_{34-71}$ with cTnI$_{147-163}$ (B), and I128 in the titration of cTnC•3Ca^{2+} with cTnI$_{128-147}$ (C), respectively, are shown in lower panel. All three residues shift from left to right. Upper panel are computer simulations using $K_D = 1$ μM and $k_{off} = 5$ s^{-1} for (A), $K_D = 150$ μM and $k_{off} = 5000$ s^{-1} for (B), and $K_D = 78$ μM and $k_{off} = 5000$ s^{-1} for (C), respectively.

D-C. $cTnI_{147-163}$ titration of $cTnC\bullet3Ca^{2+}\bullet cTnI_{34-71}$

Previously, we have titrated $cTnI_{147-163}$ to $cNTnC\bullet Ca^{2+}$ and determined a binding affinity (K_D) of 154 ± 10 μM [13]. In this study, we show the titration of this peptide to $cTnC\bullet3Ca^{2+}\bullet cTnI_{34-71}$. In this complex, the C-domain hydrophobic pocket is occupied by $cTnI_{34-71}$, which binds tightly ($K_D \leq 1$ μM). Figure D-5B shows the superimposition of the 2D $\{^1H, ^{15}N\}$ HSQC NMR spectra of $cTnC\bullet3Ca^{2+}\bullet cTnI_{34-71}$ titrated by $cTnI_{147-163}$. Similar to $cTnI_{147-163}$ binding to $cNTnC\bullet Ca^{2+}$, this reaction occurs with fast exchange kinetics on the NMR time scale. Figure D-6B shows the shifts of A31 resonance and the chemical shift changes of A31 as a function of peptide to protein ratios were fit to the equation

$$cTnC\bullet3Ca^{2+}\bullet cTnI_{34-71} + cTnI_{147-163} \leftrightarrow cTnC\bullet3Ca^{2+}\bullet cTnI_{34-71}\bullet cTnI_{147-163}$$

and yielded a K_D of 150 ± 10 μM. Within experimental error, this affinity agrees with that of $cTnI_{147-163}$ for $cNTnC\bullet Ca^{2+}$, indicating that the affinity of this peptide for the N-domain of cTnC is independent of the presence of the C-domain. The line shapes of A31 were simulated using experimentally derived values of $\Delta\nu_f$ = 21 Hz, $\Delta\nu_b$ = 16 Hz, δ_f = 3938 Hz, δ_b = 4012 Hz, $\Delta\delta$ = 74 Hz, and K_D = 150 μM and adjusting k_{off}. A k_{off} of 5000 s^{-1} provides the closest fit to the experimental data (Figure D-7B). This indicates that the dissociation rate of this region of cTnI from cTnC is in the same order of Ca^{2+} dissociation from site II and is fast enough to regulate muscle contraction.

D-D. cTnI$_{128-147}$ titration of cTnC•3Ca^{2+}

This titration was done in a previous study and we have reported the binding affinity of 78 ± 10 μM [31], but the line shape analysis was not done. For the purpose of comparing with the binding of the other two regions of cTnI to cTnC, the titration plot of cTnC•3Ca^{2+} with cTnI$_{128-147}$ is shown in Figure D-5C. Figure D-6C shows the shifts of the I128 resonance. The chemical shift changes of I128 as a function of peptide to protein ratios were fit to the equation

$$\mathrm{cTnC{\bullet}3Ca^{2+}} + \mathrm{cTnI_{128\text{-}147}} \leftrightarrow \mathrm{cTnC{\bullet}3Ca^{2+}{\bullet}cTnI_{128\text{-}147}}$$

and yielded a K_D of 78 ± 10 μM. The line shape of I128 resonances were simulated using experimentally derived values of $\Delta\nu_f = 18$ Hz, $\Delta\nu_b = 20$ Hz, $\delta_f =$ 4037 Hz, $\delta_b = 4103$ Hz, $\Delta\delta = 66$ Hz, and $K_D = 78$ μM and adjusting k_{off}. A k_{off} of 5000 s^{-1} provides the closest fit to the experimental data (Figure D-7C). This indicates that the dissociation rate of this region of cTnI from cTnC is in the same order of Ca^{2+} dissociation from site II and is fast enough to regulate muscle contraction.

DISCUSSION

The key events in regulating skeletal and cardiac muscle contraction involve Ca^{2+} binding and release from TnC, the accompanied structural changes, and the subsequent TnC-TnI interactions. Although the sequence of events involved in initiating skeletal and cardiac muscle contraction are very similar, the kinetics and thermodynamics of these events must differ for the two systems to account for the different physiological behavior of the two muscle types. The goal of this study is to characterize the kinetics of Ca^{2+} binding to intact human cTnC and of three cTnI peptides binding to the Ca^{2+}-saturated cTnC by the means of 2D {^{1}H, ^{15}N} HSQC NMR spectroscopy at 30°C, a temperature close to that of human heart (~ 37°C). The kinetics of these events are reflected in the Ca^{2+}- and peptide- induced NMR spectral intensity, chemical shift, and line-shape changes of cTnC. Analysis of the NMR spectral changes allow us to determine the dissociation (K_D) and off rate (k_{off}) constants and to derive the association rate (k_{on}) constants using the relationship $K_D = k_{off}/k_{on}$.

We first examined Ca^{2+} binding to both classes of sites in intact human cTnC. The results show that sites III and IV exhibit 100 fold higher Ca^{2+} affinity than site II ($K_{D(III,IV)}$= ~ 0.2 μM, $K_{D(II)}$= ~ 20 μM), but that site II is partially occupied before sites III and IV are saturated. The addition of the first two equivalents of Ca^{2+} saturates ~ 90% of sites III and IV and ~ 20% of site II and three equivalents saturate all three sites completely. This suggests that the Ca^{2+} occupancy of all three sites may contribute to the Ca^{2+}-dependent regulation in muscle contraction. A similar phenomenon in Ca^{2+} binding to CaM has been reported [39], demonstrating that the addition of 2 equivalents of Ca^{2+} saturates 15-35% of the N-domain sites (I/II) and 85-55% of the C-domain sites (III/IV). This is not surprising since both classes of binding sites in CaM play regulatory roles (see [40] for a review). For cTnC or sTnC, such a behavior had not been reported. Biekofsky et al. 1998 also reported line broadening of the N-domain resonances (*e.g.* I27 and I63) during Ca^{2+} binding to the C-domain and interpreted it as due to the partial occupancy of sites I/II in CaM. In this work, we also observed line broadening of the N-domain resonances (*e.g.* G70 and V72, Figure VI-2) during Ca^{2+} binding to the C-domain of cTnC. V72 is the residue equivalent to I63 in CaM. Clearly, this is the result of the partial occupancy of site II before the $[Ca^{2+}]_{total}/[cTnC]_{total}$ ratio reaches 2. Thus, no domain-domain interactions need be introduced to understand the NMR spectral changes occurring during Ca^{2+} titration of cTnC.

A second major finding from this study is that a much higher site II Ca^{2+} dissociation rate ($k_{off} = 5000\ s^{-1}$, 30°C) was observed as compared to previous published results. Such a rapid off rate had not yet been measured. A number of studies have been performed on the kinetics of Ca^{2+} binding to sTnC or cTnC, isolated or in the troponin complex and the experiments involve the use of proteins with mutated or attached fluorescent reporter groups. In sTnC, Ca^{2+} binding to sites I/II appears to be diffusion limited ($k_{on} = \sim 10^8\ M^{-1}s^{-1}$) with a Ca^{2+} dissociation rate (k_{off}) of $\sim$ 400-500 s^{-1} and the conformational change occur almost simultaneously with the Ca^{2+} association/dissociation[19-21]. In cTnC, Ca^{2+} binding to site II is also diffusion limited ($k_{on} = \sim 10^8\ M^{-1}s^{-1}$, $k_{off} = \sim 500\text{-}800\ s^{-1}$), but the conformational change was found to be significantly slower than the Ca^{2+} on- and off-rates[22-24]. These reported on- and off- rates may not, however, represent the time scales of the molecular events that would occur at physiological temperature (*e.g.* 37°C in human heart), since most of these experiments were performed at 4°C because Ca^{2+} kinetics become too fast to be measured by stopped flow methods at temperatures higher than 4°C [24]. Moreover, it is well known that both the time to peak tension after excitation and the relaxation time can vary widely with temperature, and are species-dependent. Knowing the dissociation constant $K_{D(II)} = 20\ \mu M$ and off rate constant $k_{off} = 5000\ s^{-1}$, the calculated on-rate of Ca^{2+} ($K_{D(II)} = k_{off}/k_{on}$) for site II is $2.5\times10^8\ M^{-1}s^{-1}$, indicating that Ca^{2+} binding to site II of cTnC is diffusion-controlled. Hence, at a temperature (30°C) close to the physiological level, Ca^{2+} association and dissociation from site II is rapid enough to account for the speed of cardiac muscle contraction and relaxation, which occur on the time scale of miliseconds [41].

The rapid rate of the rise of the tension necessitates a rapid binding of Ca^{2+} to the Ca^{2+}-specific regulatory site II of cTnC as well as a rapid propagation of these Ca^{2+} induced conformational changes to other components of the thin filaments, such as cTnI. Limited information is available on the kinetics of sTnC-sTnI or cTnC-cTnI interactions. In a previous study, we initiated the kinetic analysis of sTnC-sTnI interaction by studying the binding of $sTnI_{115-131}$ peptide to $sNTnC{\bullet}2Ca^{2+}$ [11]. In the present work, we have determined the off rate constants for the binding of three cTnI peptide to cTnC: these are 5 s^{-1} for $cTnI_{34-71}$, 5000 s^{-1} for $cTnI_{128-147}$, and 5000 s^{-1} for $cTnI_{147-163}$, respectively. Thus, it appears that the k_{off} for the binding of $cTnI_{128-147}$ and $cTnI_{147-163}$ to cTnC are in the same order as the k_{off} for the binding of Ca^{2+} to site II of cTnC and these events are fast enough to be kinetically competent for muscle contraction, while that of $cTnI_{34-71}$ may be too slow for this process. These three regions of cTnI have been identified by many biophysical and biological studies to be responsible for interacting with cTnC. In the antiparallel arrangement of cTnC-cTnI interaction[15, 42], $cTnI_{34-71}$, corresponding to $sTnI_{1-40}$, binds tightly to the Ca^{2+}-saturated C-domain of cTnC and the binding site was identified to be in the hydrophobic pocket of the C-domain [43]. This region of cTnI presumably adopts a similar α-helical conformation similar to that of $sTnI_{1-47}$ observed in the X-ray structure of $sTnC{\bullet}2Ca^{2+}{\bullet}sTnI_{1-47}$ complex [44]. The inhibitory domain of cTnI ($cTnI_{128-147}$) binds to the central helix area toward the C-domain of cTnC and this binding constitutes a major switch between muscle contraction and relaxation [45], and this switch is modulated by the interaction of the C-terminal region of cTnI and cTnC[46, 47]. The residues (~ 147-163) immediately following the inhibitory region bind to the hydrophobic pocket of the N-domain of cTnC. In the NMR structure of $cTnI_{147-163}$ in complex with $cNTnC{\bullet}Ca^{2+}$, $cTnI_{147-163}$ forms an α-helix and interacts with the hydrophobic surface of the N-domain stabilizing its open conformation [13]. Two models of sTnC-sTnI interactions have been proposed to rationalize the functional roles of these three regions of TnI. One proposes that the N-terminal and the inhibitory regions share overlapping binding sites on the C-domain of sTnC, which are alternatively occupied by either one or the other depending on the interactions between the N-domain of TnC and the C-domain of TnI [29]. Another proposes that the N-terminal region of TnI always binds to the C-domain of TnC regardless of the Ca^{2+}-dependent interaction between the N-domain of TnC and the C-domain of TnI, while the inhibitory region interacts with the central helix area[48, 49]. Clearly, our results are in line with the second model, which implies a structural role for the $cTnI_{34-71}$ region but functional roles for both the $cTnI_{128-147}$ and $cTnI_{147-163}$ regions.

LITERATURE CITED

1. Geeves, M.A. and K.C. Holmes, Structural mechanism of muscle contraction. Annu Rev Biochem, 1999. 68; 687-728.
2. Gordon, A.M., E. Homsher, and M. Regnier, Regulation of contraction in striated muscle. Physiol Rev, 2000. 80(2); 853-924.
3. Gagne, S.M., et al., The NMR angle on troponin C. Biochem Cell Biol, 1998. 76(2-3); 302-12.
4. van Eerd, J. and K. Takahashi, The amino acid sequence of bovine cardiac troponin-C. Comparison with rabbit skeletal troponin-C. . Biochem Biophys Res Commun, 1975. 64; 122-127.
5. Gagne, S.M., et al., Structures of the troponin C regulatory domains in the apo and calcium-saturated states. Nat Struct Biol, 1995. 2(9); 784-9.
6. Sia, S.K., et al., Structure of cardiac muscle troponin C unexpectedly reveals a closed regulatory domain. J Biol Chem, 1997. 272(29); 18216-21.
7. Spyracopoulos, L., et al., Calcium-induced structural transition in the regulatory domain of human cardiac troponin C. Biochemistry, 1997. 36(40); 12138-46.
8. Sheng, Z., et al., Isolation, expression, and mutation of a rabbit skeletal muscle cDNA clone for troponin I. The role of the NH2 terminus of fast skeletal muscle troponin I in its biological activity. J Biol Chem, 1992. 267(35); 25407-13.
9. Putkey, J.A., H.L. Sweeney, and S.T. Campbell, Site-directed mutation of the trigger calcium-binding sites in cardiac troponin C. J Biol Chem, 1989. 264(21); 12370-8.
10. Herzberg, O., J. Moult, and M.N. James, A model for the Ca2+-induced conformational transition of troponin C. A trigger for muscle contraction. J Biol Chem, 1986. 261(6); 2638-44.

11. McKay, R.T., et al., Interaction of the second binding region of troponin I with the regulatory domain of skeletal muscle troponin C as determined by NMR spectroscopy. J Biol Chem, 1997. 272(45); 28494-500.
12. McKay, R.T., et al., Defining the region of troponin-I that binds to troponin-C. Biochemistry, 1999. 38(17); 5478-89.
13. Li, M.X., L. Spyracopoulos, and B.D. Sykes, Binding of cardiac troponin-I147-163 induces a structural opening in human cardiac troponin-C. Biochemistry, 1999. 38(26); 8289-98.
14. Spyracopoulos, L., R.T. McKay, and B.D. Sykes, Structure of skeletal troponin-C in complex with skeletal troponin-I 115-131: comparison of cardiac troponin-C-troponin-I 147-163 complex. Biophysical Journal, 2000. 78; 434A.
15. Farah, C.S. and F.C. Reinach, The troponin complex and regulation of muscle contraction. Faseb J, 1995. 9(9); 755-67.
16. Solaro, R.J. and H.M. Rarick, Troponin and tropomyosin: proteins that switch on and tune in the activity of cardiac myofilaments. Circ Res, 1998. 83(5); 471-80.
17. McKay, R.T., et al., Energetics of the induced structural change in a Ca2+ regulatory protein: Ca2+ and troponin I peptide binding to the E41A mutant of the N-domain of skeletal troponin C. Biochemistry, 2000. 39(41); 12731-8.
18. Pearlstone, J.R., et al., Biological function and site II Ca2+-induced opening of the regulatory domain of skeletal troponin C are impaired by invariant site I or II Glu mutations. J Biol Chem, 2000. 275(45); 35106-15.
19. Johnson, J.D., et al., Modulation of Ca2+ exchange with the Ca(2+)-specific regulatory sites of troponin C. J Biol Chem, 1994. 269(12); 8919-23.
20. Rosenfeld, S.S. and E.W. Taylor, Kinetic studies of calcium and magnesium binding to troponin C. J Biol Chem, 1985. 260(1); 242-51.

21. Rosenfeld, S.S. and E.W. Taylor, Kinetic studies of calcium binding to regulatory complexes from skeletal muscle. J Biol Chem, 1985. 260(1); 252-61.

22. Dong, W., et al., Kinetic studies of calcium binding to the regulatory site of troponin C from cardiac muscle. J Biol Chem, 1996. 271(2); 688-94.

23. Dong, W.J., et al., A kinetic model for the binding of Ca2+ to the regulatory site of troponin from cardiac muscle. J Biol Chem, 1997. 272(31); 19229-35.

24. Hazard, A.L., et al., The kinetic cycle of cardiac troponin C: calcium binding and dissociation at site II trigger slow conformational rearrangements. Protein Sci, 1998. 7(11); 2451-9.

25. Li, M.X., et al., Calcium binding to the regulatory N-domain of skeletal muscle troponin C occurs in a stepwise manner. Biochemistry, 1995. 34(26); 8330-40.

26. Mercier, P., M.X. Li, and B.D. Sykes, Role of the structural domain of troponin C in muscle regulation: NMR studies of Ca2+ binding and subsequent interactions with regions 1-40 and 96-115 of troponin I. Biochemistry, 2000. 39(11); 2902-11.

27. Li, M.X., et al., NMR studies of Ca2+ binding to the regulatory domains of cardiac and E41A skeletal muscle troponin C reveal the importance of site I to energetics of the induced structural changes. Biochemistry, 1997. 36(41); 12519-25.

28. Wang, X., et al., Structure of the C-domain of human cardiac troponin C in complex with the Ca2+ sensitizing drug EMD 57033. J Biol Chem, 2001. 276(27); 25456-66.

29. Tripet, B., J.E. Van Eyk, and R.S. Hodges, Mapping of a second actin-tropomyosin and a second troponin C binding site within the C terminus of troponin I, and their importance in the Ca2+-dependent regulation of muscle contraction. J Mol Biol, 1997. 271(5); 728-50.

30. Smillie, L.B. and M. Nattriss, Amino acid analyses of proteins and peptides: an overview., in HPLC of Peptides and Proteins: Separation,

Analysis, and Conformation., M. C and R.S. Hodges, Editors. 1990, CRC Press: Boca Raton. p. 809-821.

31. Li, M.X., et al., Interaction of cardiac troponin C with Ca(2+) sensitizer EMD 57033 and cardiac troponin I inhibitory peptide. Biochemistry, 2000. 39(30); 8782-90.

32. Kay, L., P. Keifer, and T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc, 1992. 114; 10663-10665.

33. Zhang, O., et al., Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. J Biomol NMR, 1994. 4(6); 845-58.

34. Delaglio, F., et al., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR, 1995. 6(3); 277-93.

35. Wuthrich, K., NMR of proteins and nucleic acids. 1986, New York: John Wiley & Sons.

36. Hannon, J.D., D.A. Martyn, and A.M. Gordon, Effects of cycling and rigor crossbridges on the conformation of cardiac troponin C. Circ Res, 1992. 71(4); 984-91.

37. Holroyde, M.J., et al., The calcium and magnesium binding sites on cardiac troponin and their role in the regulation of myofibrillar adenosine triphosphatase. J Biol Chem, 1980. 255(24); 11688-93.

38. Johnson, J.D., et al., A fluorescent probe study of Ca2+ binding to the Ca2+-specific sites of cardiac troponin and troponin C. J Biol Chem, 1980. 255(20); 9635-40.

39. Biekofsky, R.R., et al., Ca2+ coordination to backbone carbonyl oxygen atoms in calmodulin and other EF-hand proteins: 15N chemical shifts as probes for monitoring individual-site Ca2+ coordination. Biochemistry, 1998. 37(20); 7617-29.

40. Crivici, A. and M. Ikura, Molecular and structural basis of target recognition by calmodulin. Annu Rev Biophys Biomol Struct, 1995. 24; 85-116.

41. Mason, E., Human Physiology, The Benjamin/Cummings series in the life sciences. 1983.

42. Krudy, G.A., et al., NMR studies delineating spatial relationships within the cardiac troponin I-troponin C complex. J Biol Chem, 1994. 269(38); 23731-5.

43. Gasmi-Seabrook, G.M., et al., Solution structures of the C-terminal domain of cardiac troponin C free and bound to the N-terminal domain of cardiac troponin I. Biochemistry, 1999. 38(26); 8313-22.

44. Vassylyev, D.G., et al., Crystal structure of troponin C in complex with troponin I fragment at 2.3-A resolution. Proc Natl Acad Sci U S A, 1998. 95(9); 4847-52.

45. Van Eyk, J.E. and R.S. Hodges, The biological importance of each amino acid residue of the troponin I inhibitory sequence 104-115 in the interaction with troponin C and tropomyosin-actin. J Biol Chem, 1988. 263(4); 1726-32.

46. Ramos, C.H., Mapping subdomains in the C-terminal region of troponin I involved in its binding to troponin C and to thin filament. J Biol Chem, 1999. 274(26); 18189-95.

47. Rarick, H.M., et al., The C terminus of cardiac troponin I is essential for full inhibitory activity and Ca2+ sensitivity of rat myofibrils. J Biol Chem, 1997. 272(43); 26887-92.

48. Luo, Y., et al., Photocrosslinking of benzophenone-labeled single cysteine troponin I mutants to other thin filament proteins. J Mol Biol, 2000. 296(3); 899-910.

49. Tung, C.S., et al., A model of troponin-I in complex with troponin-C using hybrid experimental data: the inhibitory region is a beta-hairpin. Protein Sci, 2000. 9(7); 1312-26.

APPENDIX E

Database

OVERVIEW

Typically NMR research laboratories are similar to a basic chemistry lab. In the Sykes lab there is some microbiology equipment for the expression, isolation, and purification of peptides and proteins. Samples are prepared by dissolving pure protein powder in a standard buffer and then a number of 1D, 2D, and 3D experiments are conducted over a few weeks. For metabolomic investigations the focus is no longer on one sample with multiple experiments, but rather one NMR experiment used to acquire spectra from hundreds of samples. *This appendix provides some insight to the administrative work I accomplished in order to handle the unique requirements of large clinical metabolomic investigations.*

INTRODUCTION

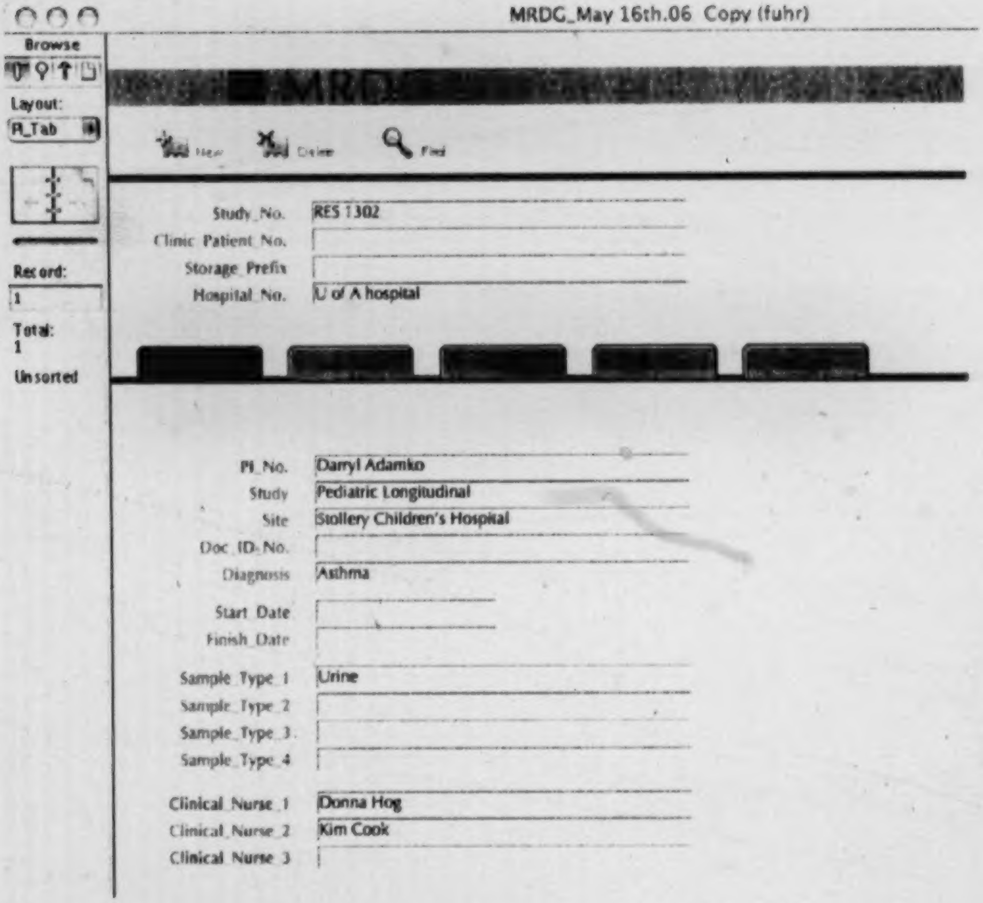
Having decided that my PhD. focus would be on clinical metabolomics of cardiac and pulmonary disorders it was clear that the Sykes laboratory located in the Medical Sciences Building would not be sufficient for the types and number of samples I would be accumulating. I moved to the Canadian National High Field NMR Centre (NANUC) to set up the spare wet lab as a biohazard laboratory. The appropriate equipment was ordered, the required documentation was prepared, the training of personnel was conducted, and the certification of the space was acquired.

Once the physical space was created the need arose to properly organize the growing amount of information and samples. Human clinical investigations have a need to handle samples and information in an ethical manner, ensuring only key individuals are privy to secure data, while allowing access to sample preparation and storage information for the technicians. The business manager of NANUC, Bruce Lix, and I developed a database system with the software File Maker Pro version 8.3 (Santa Clara, CA.). It was necessary that the database have the capacity to manage the growing number of collaborations, each with multiple on-going studies, the hundreds of samples associated with each study, and the pertinent information regarding sample receiving, preparation, acquisition, and analysis. In addition, for human studies there is a large amount of clinical information that is associated with each sample.

Below are some examples of the different processes that had to be captured and documented in the database. In the end a bar-coding system was integrated into the database so that each sample could be assigned a barcode. With this system the technician could quickly scan the top of the Eppendorff or NMR tube and automatically retrieve all the relevant information about the particular sample.

E.1Patient Information

Within the clinical portion of the database it was decided to organize the patients principally by four items; Study Number (the ethics number), Clinical Patient Number (patient ID number from the hospital), Storage Prefix (a linked number for sample retrieval in NANUC), and Hospital Number (the number of the hospital where admitted). Next is a *Private Investigator* (PI) section, which was established to distinguish the particular study the patient is participating in (often a single PI has multiple studies).


MRDG_May 16th.06 Copy (fuhr)
Browse
Layout:
PI_Tab
MRD
New
Delete
Find
Record:
1
Total:
1
Unsorted

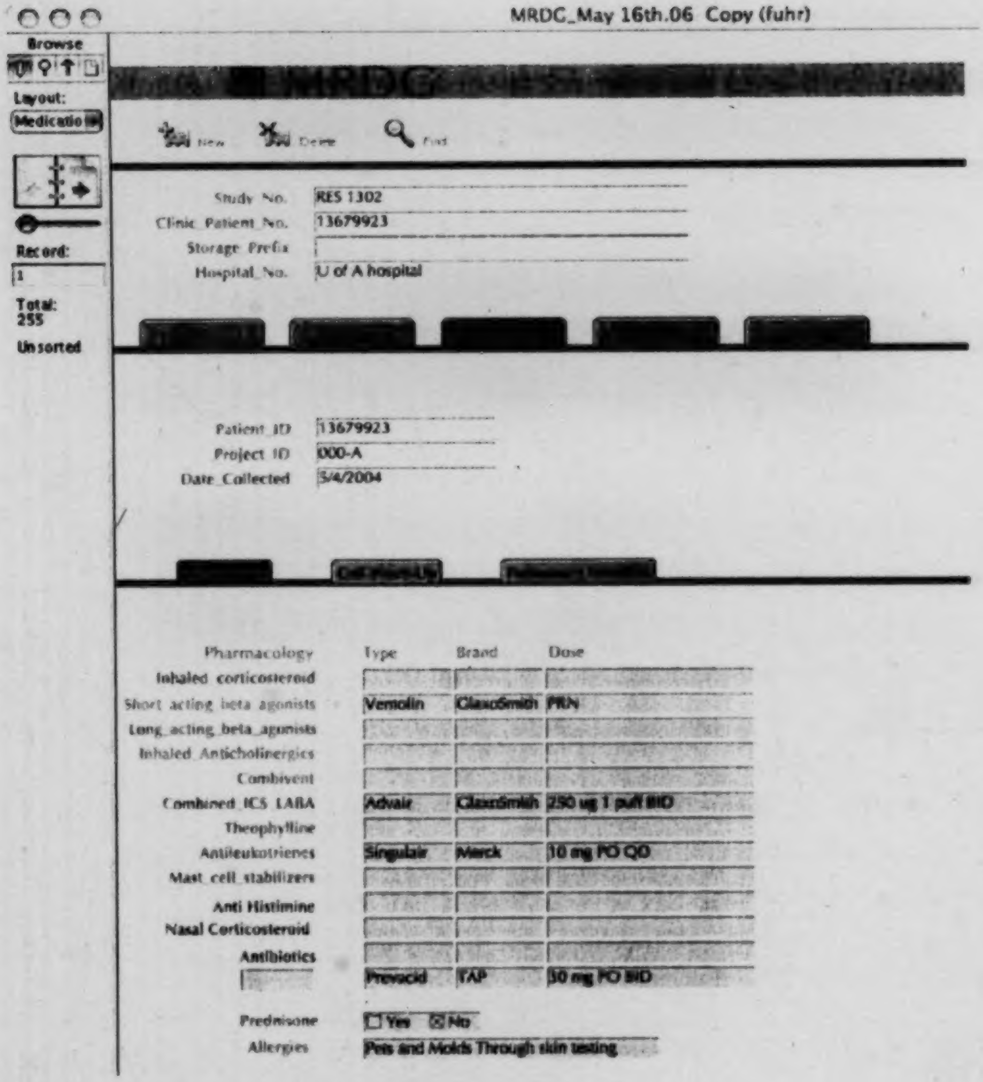
Study_No. RES 1302
Clinic_Patient_No.
Storage_Prefix
Hospital_No. U of A hospital

PI_No. Darryl Adamko
Study Pediatric Longitudinal
Site Stollery Children's Hospital
Doc_ID_No.
Diagnosis Asthma
Start_Date
Finish_Date
Sample_Type_1 Urine
Sample_Type_2
Sample_Type_3
Sample_Type_4
Clinical_Nurse_1 Donna Hog
Clinical_Nurse_2 Kim Cook
Clinical_Nurse_3

Figure E-1. **Principle investigator section.**

Under the *Individual* tab a number of fields can be completed regarding the patient's: name, place of residence, health care number, DOB, age, gender, and ethnicity. This tab is particularly important for epidemiological studies and was useful when reviewing the ethnicity of the normal subjects in Chapter IV. For privacy purposes I will not show the *Individual* screen here.

By selecting the *Patient* tab three fields appear as well as three sub-screens that may be chosen. The three fields are Patient ID (the same as the *Clinical Patient* number from above), *Project ID* (the separate number assigned to the patient upon entering the study), and *Date Collected*. The first of the three sub-screens that may be selected is titled medication and is where all of the prescribed medications are filled in (*e.g.* inhaled corticosteroids, short acting beta agonists, anticholinergics, theophylline, leukotriene receptor antagonists, anti-histamine, antibiotics), as well as any allergies the patient may have. The second sub-screen is *Cell Work-up* and provides fields to be filled in following a BAL; total cell count, viability, neutrophils, eosinophils, macrophages, lymphocytes, and epithelial cells. Finally, the last sub-screen is identified as *Pulmonary Function*. Here data such as atopy, forced expiratory volume (FEV1), forced vital capacity (FVC), FEV1/FVC, forced expiratory flow (FEF 25-75%), and peak expiratory flow (PEF) can be entered.



MRDC_May 16th.06 Copy (fuhr)
Browse
Layout:
Medicatio
Record:
1
Total:
255
Unsorted
MRDC
New
Delete
Find

Study_No. RES 1302

Clinic_Patient_No. 13679923

Storage_Prefix

Hospital_No. U of A hospital

Patient_ID 13679923

Project_ID 000-A

Date_Collected 5/4/2004

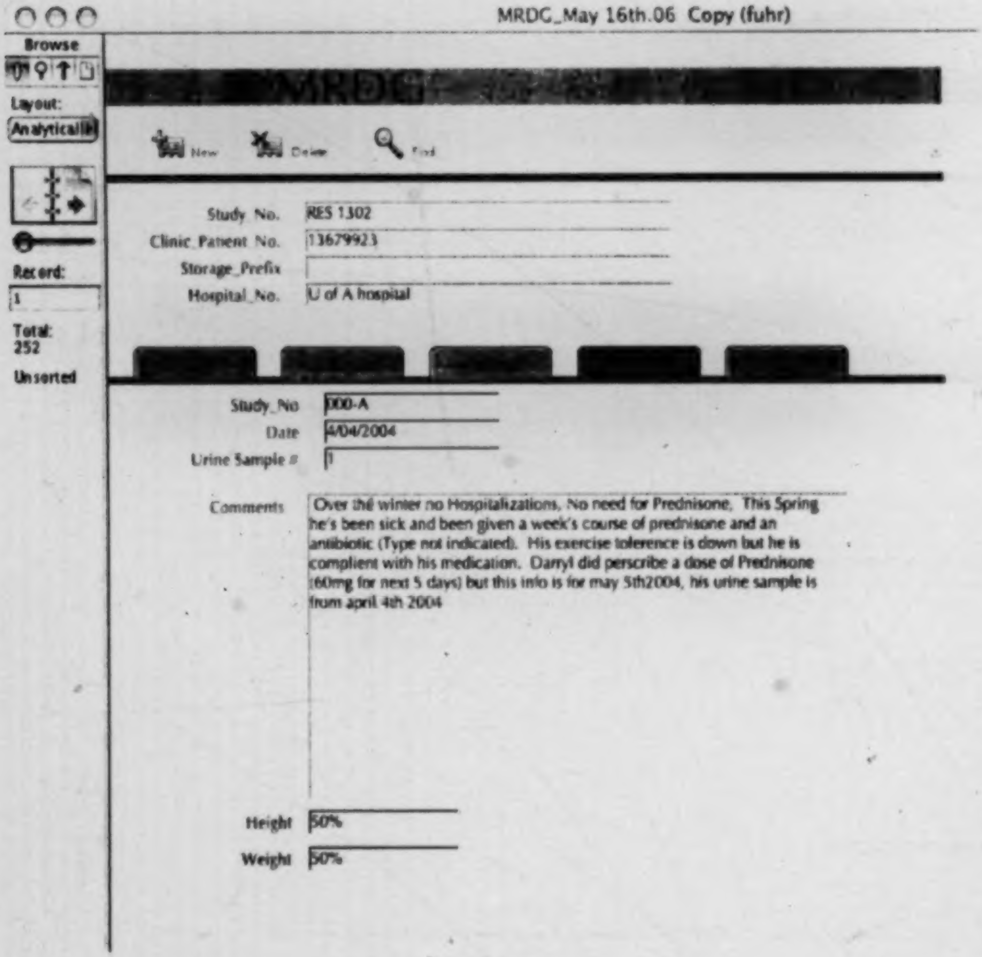
Pharmacology	Type	Brand	Dose
Inhaled_corticosteroid			
Short_acting_beta_agonists	Vemolin	GlaxoSmith	PRN
Long_acting_beta_agonists			
Inhaled_Anticholinergics			
Combivent			
Combined_ICS_LABA	Advair	GlaxoSmith	250 ug 1 puff BID
Theophylline			
Antileukotrienes	Singulair	Merck	10 mg PO QD
Mast_cell_stabilizers			
Anti Histimine			
Nasal Corticosteroid			
Antibiotics			
	Prevacid	TAP	30 mg PO BID

Prednisone ☐ Yes ☒ No

Allergies Pets and Molds Through skin testing

Figure E-2. **Patient clinical information section.**

The *Analytical* tab provides a textbox for the addition of comments that may be pertinent to the care of the patient and the study. There are also two fields for the documentation of the patient's height and weight progress.



Record: 1

Total: 252

Study_No. RES 1302

Clinic_Patient_No. 13679923

Storage_Prefix

Hospital_No. U of A hospital

Study_No 000-A

Date 4/04/2004

Urine Sample # 1

Comments Over thé winter no Hospitalizations, No need for Prednisone. This Spring he's been sick and been given a week's course of prednisone and an antibiotic (Type not indicated). His exercise tolerence is down but he is complient with his medication. Darryl did perscribe a dose of Prednisone (60mg for next 5 days) but this info is for may 5th2004, his urine sample is from april 4th 2004

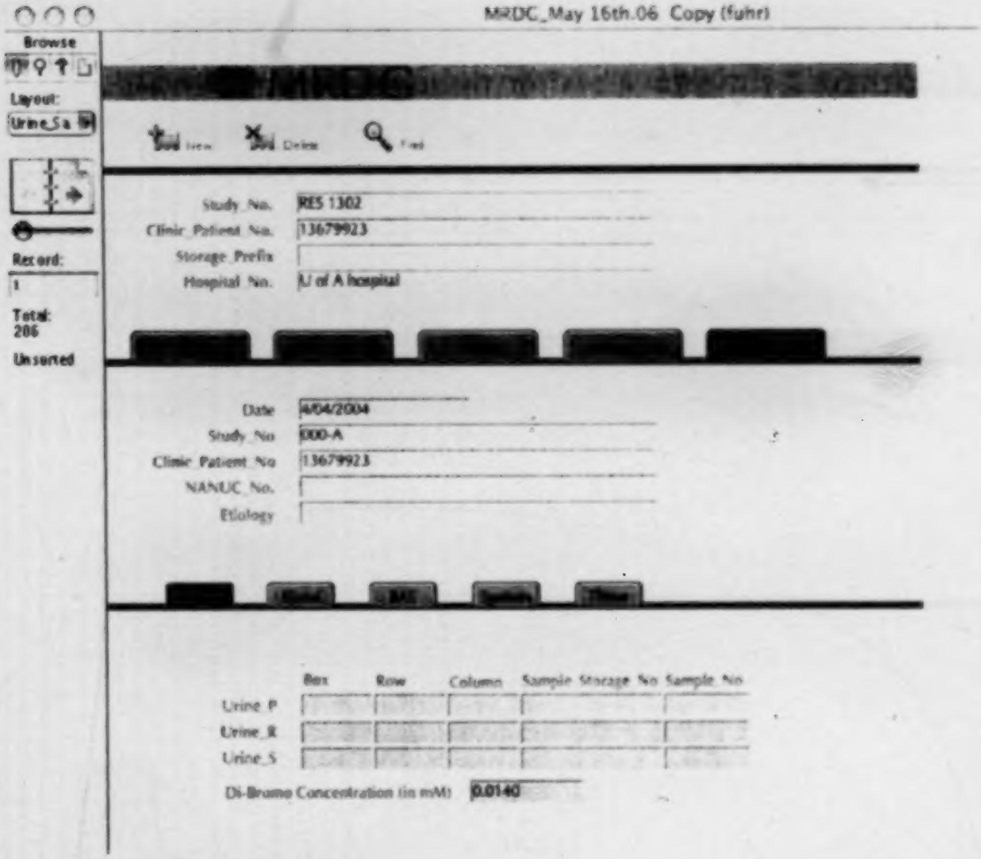
Height 50%

Weight 50%

Figure E-3. **Analytical textbox**.

Finally the *Sample* tab opens a sub-region with five fields, the unique fields include; Study No (the study number is the patients research project number), NANUC number (is the sequential number given to all samples received by NANUC), and finally Etiology (used primarily for the pneumonia project).

Within the *Sample* tab are five additional sub-category tabs. These include urine, blood, BAL, sputum, and tissue. These sample types were selected, as they were the primary biological sampling methods I employed in my investigations. Additional tabs may be added at any time. Within each sample type is a field, which allows for the documentation of where the sample is physically located in NANUC. For urine there are three samples always on hand. 'P' is the pH'd urine sample (700µl), 'R' is the raw and unhandled sample (700µl), and 'S' is the 3 ml untouched stored sample. Each sample has a particular cardboard box that it is stored in and placed within the deep-freeze. Within the box the sample has a designated row and column. The Sample Storage number and Sample number are additional cross-referencing tools. The new bar coding system allows for automated retrieval of this information by simply scanning the top of the Eppendorff, or sample tube.



Record: 1

Total: 206

Study_No. RES 1302

Clinic_Patient_No. 13679923

Storage_Prefix

Hospital_No. U of A hospital

Date 4/04/2004

Study_No 000-A

Clinic_Patient_No 13679923

NANUC_No.

Etiology

	Box	Row	Column	Sample_Storage_No	Sample_No
Urine_P					
Urine_R					
Urine_S					

Di-Bromo Concentration (in mM) 0.0140

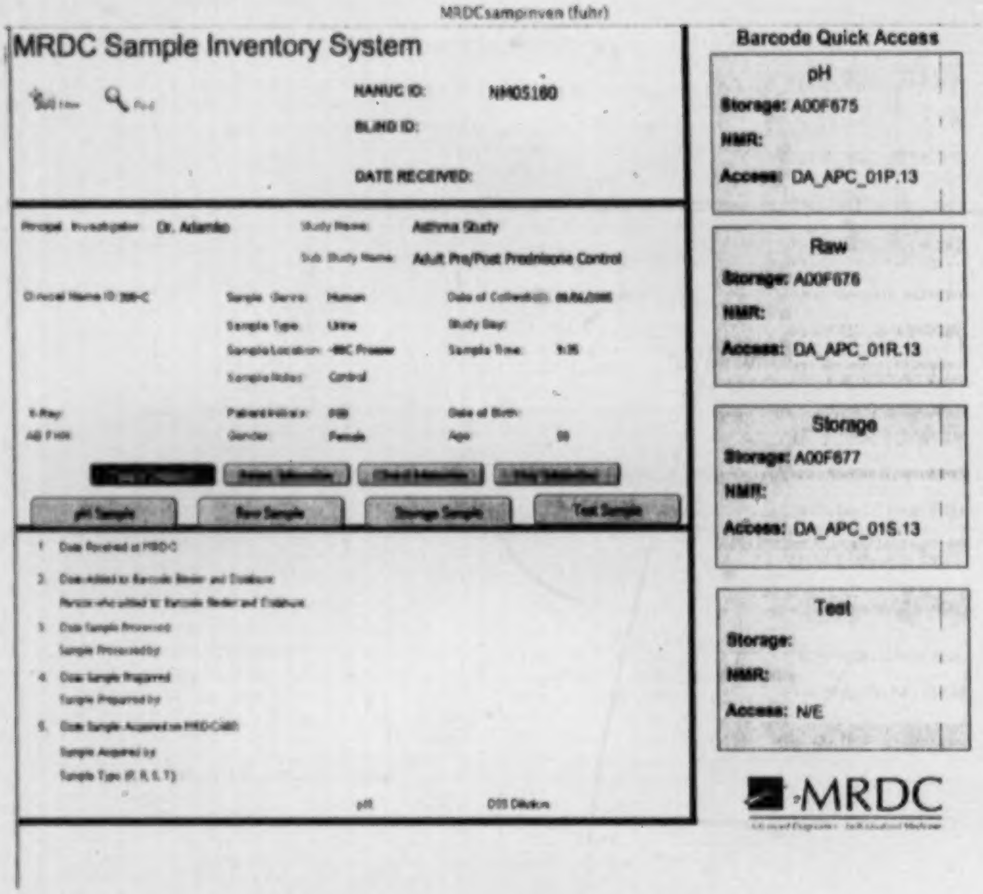
Figure E-4. **Sample storage page.**

E.2 Sample Inventory

When I began my PhD. research an organizational platform needed to be created due to my multiple collaborations, each with many ongoing research projects. A group called MRDG (magnetic resonance diagnostic group) was formed and served as a larger collaborative organization through which ideas and resources could be pooled. This grew into what is now called MRDC (magnetic resonance diagnostic centre) and includes research projects outside my immediate involvement. However, with so many samples arriving to NANUC from countless locations (different countries, provinces, and hospitals) and for multiple projects the sample inventory database was created.

The sample inventory database presents a Barcode Quick Access column on the right of the screen that presents the Storage and Access information for the sample. In the main page the basic information is depicted; NANUC ID number, blinded ID, date received, PI, study name, sub-study name, clinical ID, gender, location, type, time of collection, initials, DOB, and age.

The first row of tabs begins with *Sample Summary*. This tab opens a text window that documents a general overview of the sample arrival to NANUC, bar coding, preparation, and acquisition.



MRDCsampinven (fuhr)

MRDC Sample Inventory System

NANUC ID: NM05180

BLIND ID:

DATE RECEIVED:

Dr. Adamko

Asthma Study

Adult Pre/Post Prednisone Control

Urine

Control

Female

Barcode Quick Access

pH
Storage: A00F675
NMR:
Access: DA_APC_01P.13

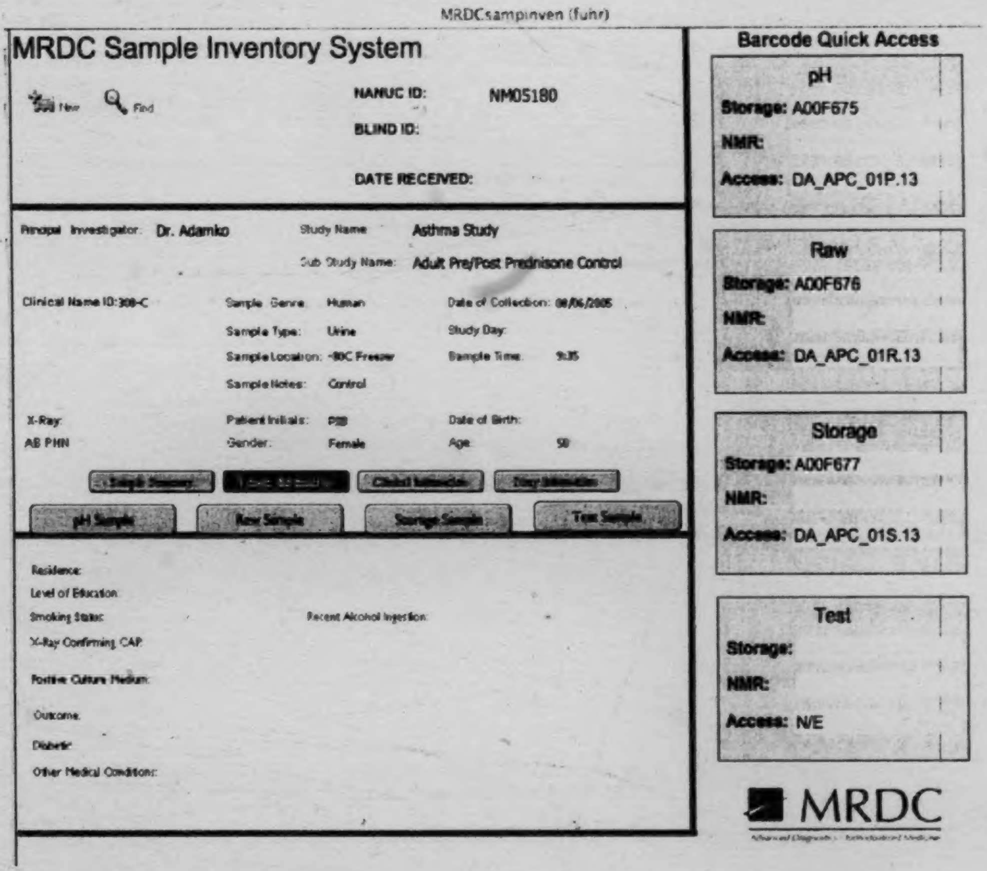
Raw
Storage: A00F676
NMR:
Access: DA_APC_01R.13

Storage
Storage: A00F677
NMR:
Access: DA_APC_01S.13

Test
Storage:
NMR:
Access: N/E

Figure E-5. **Sample inventory main page and sample summary.**

The second tab is termed *Patient Information* and retrieves a simple text window with basic patient information such as smoking status, diabetes, alcohol consumption, and level of education.


MRDCsampinven (fuhr)
New
Find
Sample Summary
Clinical Information
pH Sample
Raw Sample
Storage Sample
Test Sample
MRDC

MRDC Sample Inventory System

NANUC ID: NM05180
BLIND ID:
DATE RECEIVED:

Principal Investigator: Dr. Adamko
Study Name: Asthma Study
Sub Study Name: Adult Pre/Post Prednisone Control
Clinical Name ID: 308-C
Sample Genre: Human
Date of Collection: 08/06/2005
Sample Type: Urine
Study Day:
Sample Location: -80C Freezer
Sample Time: 9:35
Sample Notes: Control
X-Ray:
AB PHN
Patient Initials: PJB
Date of Birth:
Gender: Female
Age: 58

Residence:
Level of Education:
Smoking Status:
Recent Alcohol Ingestion:
X-Ray Confirming CAP:
Positive Culture Medium:
Outcome:
Diabetic:
Other Medical Conditions:

Barcode Quick Access

pH
Storage: A00F675
NMR:
Access: DA_APC_01P.13

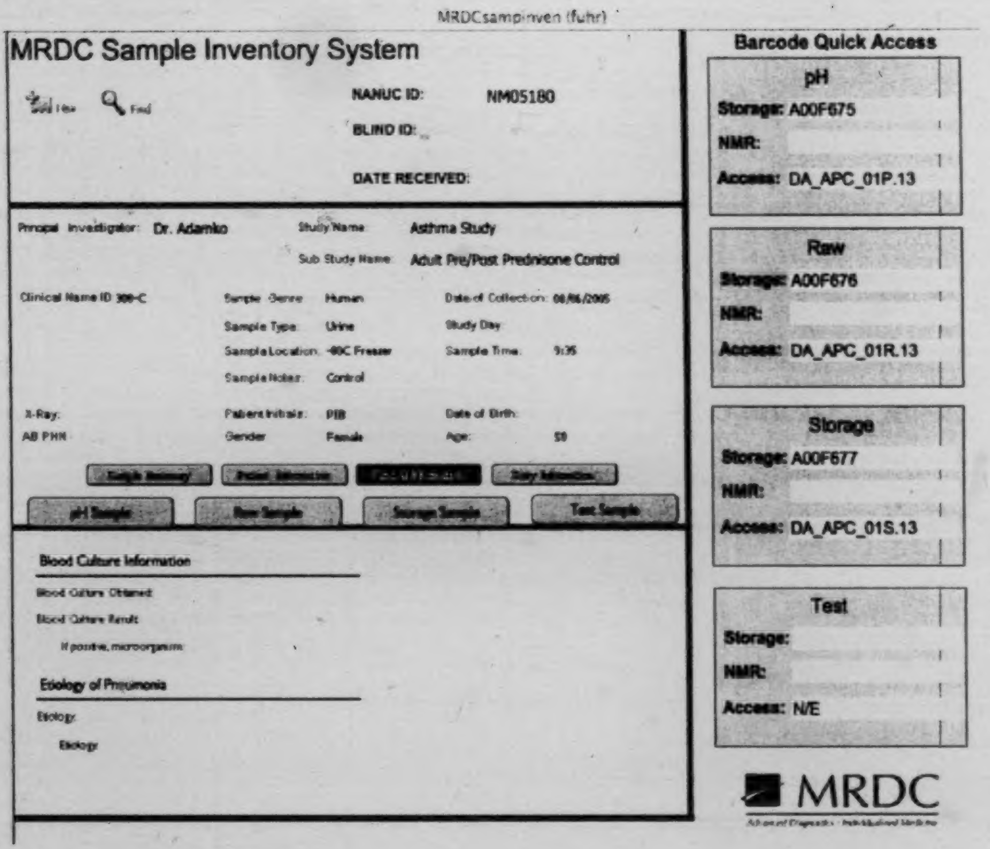
Raw
Storage: A00F676
NMR:
Access: DA_APC_01R.13

Storage
Storage: A00F677
NMR:
Access: DA_APC_01S.13

Test
Storage:
NMR:
Access: N/E

Figure E-6. **Patient information**.

The next tab is Clinical Information and is primarily concerned with an ongoing study of pneumonia patients and associated etiology.


MRDCsampinven (fuhr)
Add New
Find
Sample Summary
Patient Information
Diary Information
pH Sample
Raw Sample
Storage Sample
Test Sample
MRDC

MRDC Sample Inventory System

NANUC ID: NM05180
BLIND ID:
DATE RECEIVED:

Principal Investigator: Dr. Adamko
Study Name: Asthma Study
Sub Study Name: Adult Pre/Post Prednisone Control
Clinical Name ID 300-C
Sample Genre: Human
Sample Type: Urine
Sample Location: -80C Freezer
Sample Notes: Control
Date of Collection: 08/06/2005
Study Day:
Sample Time: 9:35
X-Ray:
AB PHN:
Patient Initials: PIB
Gender: Female
Date of Birth:
Age: 59

Blood Culture Information
Blood Culture Obtained:
Blood Culture Result:
If positive, microorganism:

Etiology of Pneumonia
Etiology:
Etiology:

Barcode Quick Access

pH
Storage: A00F675
NMR:
Access: DA_APC_01P.13

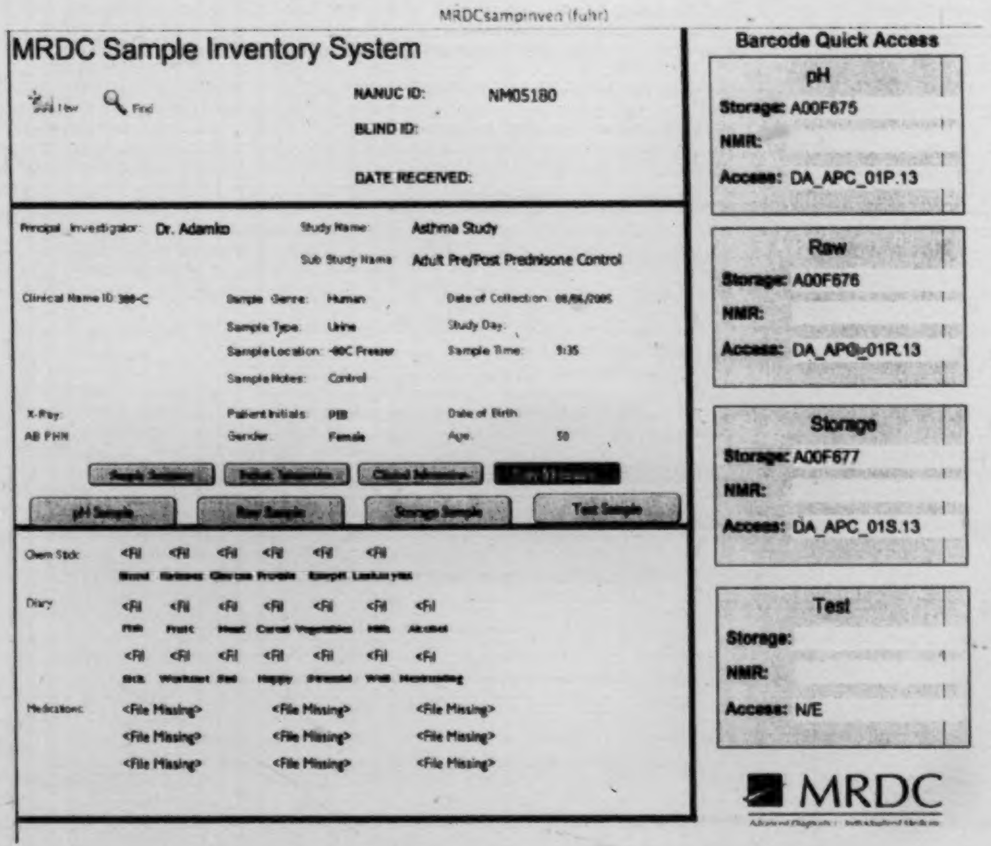
Raw
Storage: A00F676
NMR:
Access: DA_APC_01R.13

Storage
Storage: A00F677
NMR:
Access: DA_APC_01S.13

Test
Storage:
NMR:
Access: N/E

Figure E-7. **Clinical information.**

The final tab in this series is Diary Information. This tab retrieves all the meta information catalogued during the normals study (see Chapter IV). This additional information includes a ChemStick (blood, ketones, glucose, etc), a diary outlining the diet, stress, and exercise of the individual, as well as all the medication taken during the 30-day study.



MRDC Sample Inventory System

NANUC ID: NM05180
BLIND ID:
DATE RECEIVED:

Principal Investigator: Dr. Adamko
Study Name: Asthma Study
Sub Study Name: Adult Pre/Post Prednisone Control
Clinical Name ID: 388-C
Sample Genre: Human
Date of Collection: 08/06/2005
Sample Type: Urine
Study Day:
Sample Location: -80C Freezer
Sample Time: 9:35
Sample Notes: Control
X-Ray:
AB PHN
Patient Initials: PIB
Date of Birth:
Gender: Female
Age: 50

Chem Stick: <Fil <Fil <Fil <Fil <Fil <Fil
Diary: <Fil <Fil <Fil <Fil <Fil <Fil <Fil
<Fil <Fil <Fil <Fil <Fil <Fil <Fil
Medications: <File Missing> <File Missing> <File Missing>
<File Missing> <File Missing> <File Missing>
<File Missing> <File Missing> <File Missing>

Barcode Quick Access

pH
Storage: A00F675
NMR:
Access: DA_APC_01P.13

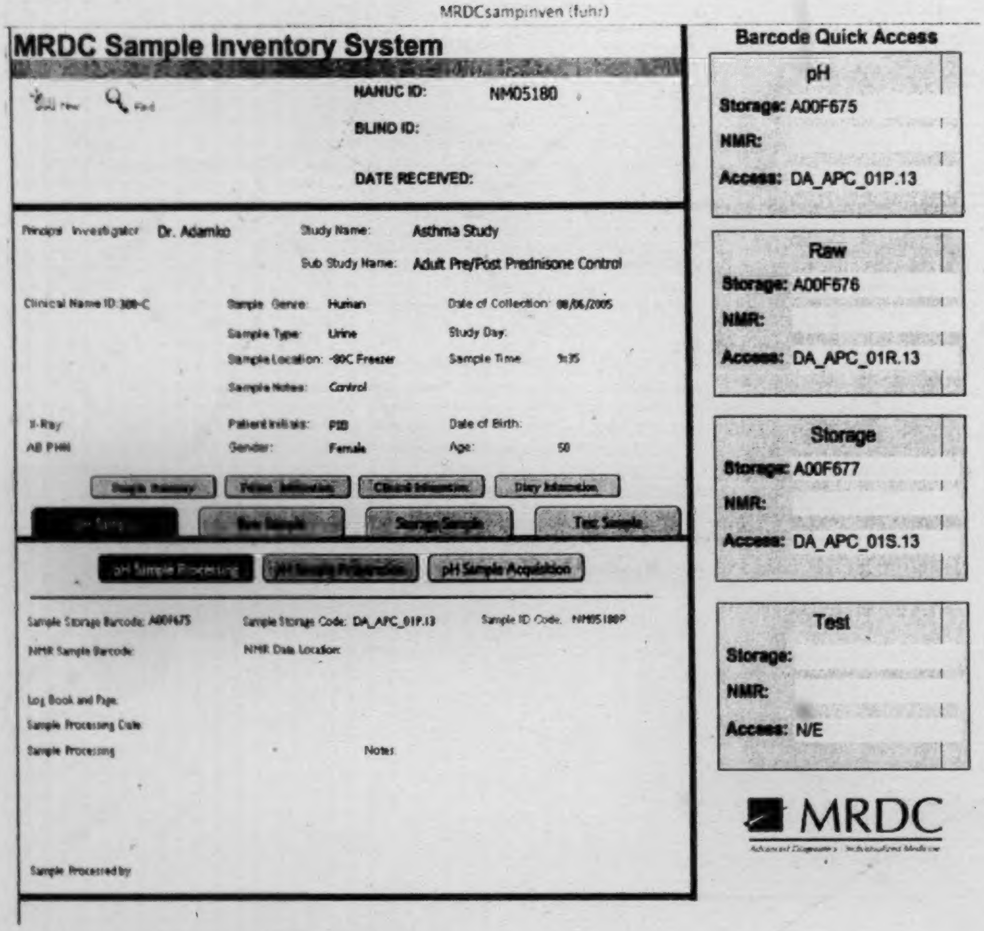
Raw
Storage: A00F676
NMR:
Access: DA_APC_01R.13

Storage
Storage: A00F677
NMR:
Access: DA_APC_01S.13

Test
Storage:
NMR:
Access: N/E

Figure E-8. **Diary information.**

Selection of one of the Sample tabs calls on another three sub-categories for selection. As an example I will select *pH Sample*, and within that page I will select *pH Sample Processing*. This window retrieves all the information regarding the processing of the sample, *e.g.* the storage barcode, storage code, sample ID code, page number in the log book, date, and who processed the samples.


MRDCsampinven (fuhr)
MRDC Sample Inventory System
New
Find
Sample Summary
Clinical Information
Diary Information
Test Sample
pH Sample Processing
pH Sample Acquisition
Barcode Quick Access
MRDC

NANUC ID: NM05180
BLIND ID:
DATE RECEIVED:

Principal Investigator: Dr. Adamko
Study Name: Asthma Study
Sub Study Name: Adult Pre/Post Prednisone Control
Clinical Name ID: 300-C
Sample Genre: Human
Date of Collection: 08/06/2005
Sample Type: Urine
Study Day:
Sample Location: -80C Freezer
Sample Time: 9:35
Sample Notes: Control
X-Ray
AB PHN
Patient Initials: PIB
Date of Birth:
Gender: Female
Age: 50

Sample Storage Barcode: A00F675
Sample Storage Code: DA_APC_01P.13
Sample ID Code: NM05180P
NMR Sample Barcode:
NMR Data Location:
Log Book and Page:
Sample Processing Date:
Sample Processing
Notes:
Sample Processed by

pH
Storage: A00F675
NMR:
Access: DA_APC_01P.13

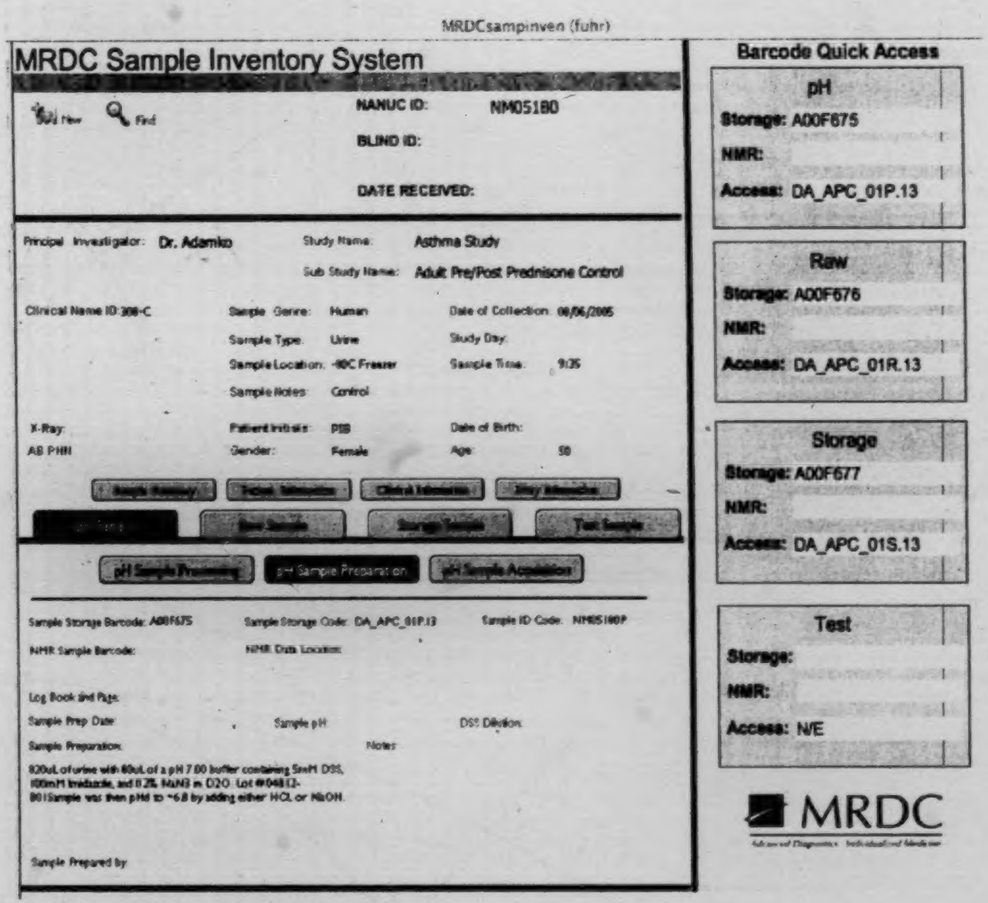
Raw
Storage: A00F676
NMR:
Access: DA_APC_01R.13

Storage
Storage: A00F677
NMR:
Access: DA_APC_01S.13

Test
Storage:
NMR:
Access: N/E

Figure E-9. **pH sample processing.**

The next tab called *pH Sample Preparation* provides all of the storage and processing information, as well as how the sample was prepared. This information includes the sample volume, pH, and what was added to the sample during preparation.


MRDCsampinven (fuhr)
MRDC Sample Inventory System
Add New
Find
Sample Summary
Patient Information
Clinical Information
Study Information
pH Sample
Raw Sample
Storage Sample
Test Sample
pH Sample Processing
pH Sample Preparation
pH Sample Acquisition
Barcode Quick Access
MRDC

NANUC ID: NM05180
BLIND ID:
DATE RECEIVED:

Principal Investigator: Dr. Adamko
Study Name: Asthma Study
Sub Study Name: Adult Pre/Post Prednisone Control
Clinical Name ID: 308-C
Sample Genre: Human
Date of Collection: 08/06/2005
Sample Type: Urine
Study Day:
Sample Location: -80C Freezer
Sample Time: 9:35
Sample Notes: Control
X-Ray:
AB PHN
Patient Initials: PSB
Date of Birth:
Gender: Female
Age: 50

Sample Storage Barcode: A00F675
Sample Storage Code: DA_APC_01P.13
Sample ID Code: NM05180P
NMR Sample Barcode:
NMR Data Location:
Log Book and Page:
Sample Prep Date:
Sample pH:
DSS Dilution:
Sample Preparation:
Notes:
820uL of urine with 80uL of a pH 7.00 buffer containing 5mM DSS, 100mM Imidazole, and 0.2% NaN3 in D2O. Lot #04812-001Sample was then pHd to ~6.8 by adding either HCL or NaOH.
Sample Prepared by

pH
Storage: A00F675
NMR:
Access: DA_APC_01P.13

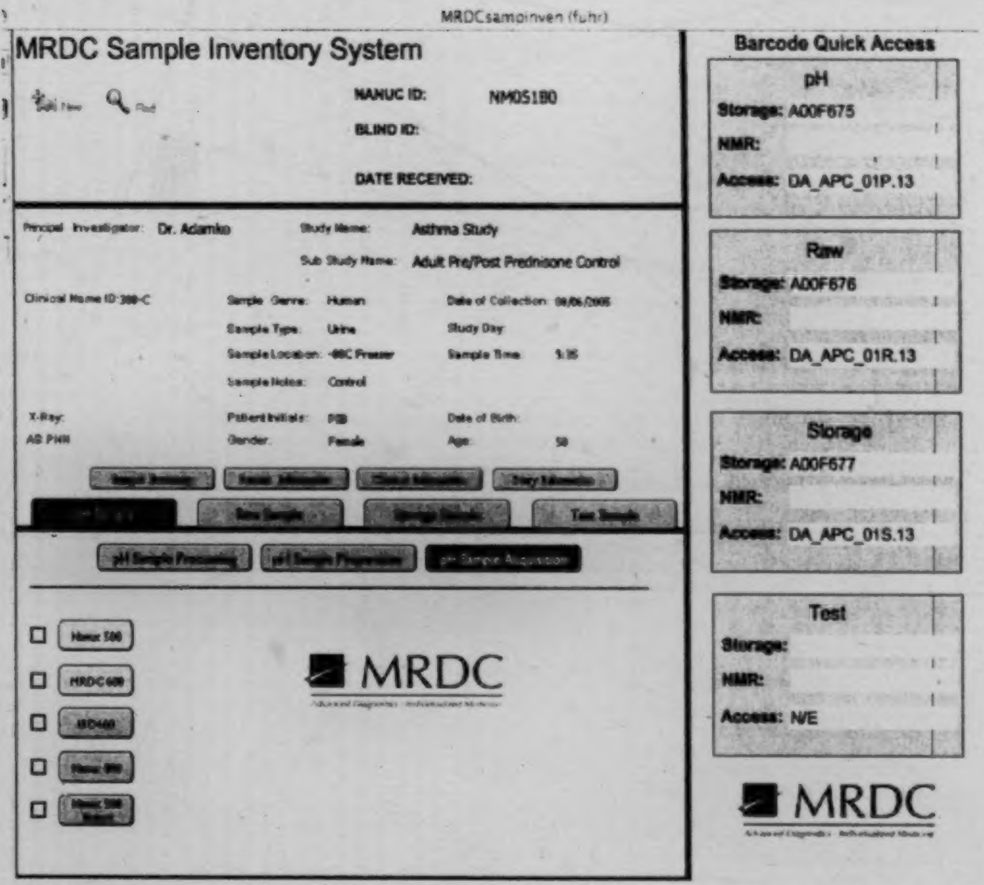
Raw
Storage: A00F676
NMR:
Access: DA_APC_01R.13

Storage
Storage: A00F677
NMR:
Access: DA_APC_01S.13

Test
Storage:
NMR:
Access: N/E

Figure E-10. **pH sample preparation**

The final tab is the *pH Sample Acquisition* tab. This tab allows the researcher to quickly view the particular spectrometer the data was collected. By clicking on the spectrometer the textbox fills with the acquisition date, name of the FID, location of the FID, the FID barcode, and the DSS linewidth.


MRDCsampinven (fuhr)
New
Find
pH Sample Processing
pH Sample Preparation
pH Sample Acquisition
Raw Sample
Storage Sample
Test Sample
MRDC

MRDC Sample Inventory System

NAHUC ID: NM05180
BLIND ID:
DATE RECEIVED:

Principal Investigator: Dr. Adamko
Study Name: Asthma Study
Sub Study Name: Adult Pre/Post Prednisone Control
Clinical Name ID: 388-C
Sample Genre: Human
Date of Collection: 08/06/2005
Sample Type: Urine
Study Day:
Sample Location: -80C Freezer
Sample Time: 9:35
Sample Notes: Control
X-Ray:
AB PHN
Patient Initials: PB
Date of Birth:
Gender: Female
Age: 58

Barcode Quick Access

pH
Storage: A00F675
NMR:
Access: DA_APC_01P.13

Raw
Storage: A00F676
NMR:
Access: DA_APC_01R.13

Storage
Storage: A00F677
NMR:
Access: DA_APC_01S.13

Test
Storage:
NMR:
Access: N/E

Figure E-11. **pH sample acquisition**

CONCLUSION

At first glance sample receiving, handling, and data acquisition seem relatively simple aspects of a research project. However, as collaborations grow and research projects expand the amount of data and samples that need to be catalogued carefully grows exponentially. This is a particular concern for clinical metabolomic studies because the research requires the collection of hundreds of samples, stored in duplicate or triplicate, each with individualized meta-information (*e.g.* patient/clinical information, diagnosis, and medications). Therefore, it is extremely important that the samples and information is sorted and linked appropriately to ensure that all of the data is correct.

Human clinical and metabolomic studies have the added pressure of the ethical treatment of information and biological samples. This involves password protecting and backing-up all information in a secure fashion. Samples are often double or triple blinded to ensure the privacy and anonymity of the research participants and the bias of the researcher. Together with Bruce Lix, Chris Skappak, Kathryn Rankin, and Shana Regush this database was created to facilitate the constant influx of samples and information, as well as to enable the continuing clinical and metabolomic research in NANUC.